

A nut to crack a sledgehammer

Antidoping scientists are failing in their efforts to rid sport of performance-enhancing drugs. But there are no inherent technical obstacles — the problem is simply one of resources.

Two years ago, Juan Antonio Samaranch, president of the International Olympic Committee (IOC), provoked a storm by suggesting in an interview that athletes should be allowed to take performance-enhancing drugs, provided the substances are not damaging to health. Samaranch subsequently claimed to have been misinterpreted, and the idea was swiftly forgotten.

On the eve of the Sydney Olympics, and with the media buzzing with suggestions that drug use in sport is rife, it is tempting to revisit this idea. If the dope testers cannot keep pace with the drug abusers, surely it would be best to concentrate on protecting athletes from themselves, rather than chasing the elusive ideal of drug-free competition? But although despair about the failure of current doping control measures is valid (see page 124), this logic is based on two erroneous assumptions.

The first is that science will never be able to solve the problem. In fact, the scientific challenges are not especially large — the main obstacle is the sums of money required to develop and validate the necessary tests. The second is that there is any meaningful distinction between 'performance-enhancing' and 'harmful' drugs — the vast majority of substances on the IOC's banned list are, to a greater or lesser degree, potentially harmful to health.

In any case, athletes seem no more likely to comply with a system that is nominally there to protect their health than with one that seeks to disqualify the 'cheats'. In a 1995 survey of elite US athletes, virtually all respondents said they would take a banned substance if they were certain not to be caught and certain to win. More than half said they would do the same if they were certain to win every competition for five years and subsequently die from the drug's side effects.

Given athletes' will to win, apparently at any personal cost, the real

problem is that the scientists appointed by sports federations to develop antidoping tests and put them into action are being given a nut to crack a sledgehammer. The solution is not to back off from testing, or to change the criteria by which drugs are put on the banned list, but to encourage sports governing bodies to open their coffers wide enough to give the scientists the resources they need to do the job effectively.

A report published last week by the US National Commission on Sports and Substance Abuse suggests that a five-year research programme of US\$100 million would put the antidoping scientists back on level terms. The report also criticizes the IOC for failing to promote the development of more new tests, and argues that only a truly independent international body — one that does not depend for its existence on the financial success of the Olympic Games — would have the necessary distance from commercial pressures to ensure that a true drug-free policy is enforced.

In theory, such a body has already been created, at least in name. The World Anti-Doping Agency (WADA), based in Lausanne, Switzerland, was established in January. For its first two years, WADA will be funded by the IOC, but thereafter should receive 50% of its money from governments. It will take over Olympic doping control after Sydney, and also work with national sporting federations to oversee out-of-competition testing.

For WADA to make any difference, it must demonstrate its independence from the IOC, gain many more affiliate countries than the handful that have so far signed up, and convince them to fund an antidoping budget that approaches the figure suggested by the US National Commission on Sports and Substance Abuse. If it fails to do this, many athletes will continue to make the choice between doping and not winning. ■

Time for a truce?

Public bickering over a controversial theory on the origins of AIDS is generating more heat than light.

By underlining the potential dangers involved in creating novel pathways by which pathogens can pass from animals to humans, the author Edward Hooper has performed a useful service. In his book *The River*, Hooper argues that the origins of HIV lie in a contaminated polio vaccine produced from chimpanzee tissue and used in Africa in the late 1950s. Even those who dismiss this argument will accept that there has been complacency about the ease with which pathogens can cross the species barrier. The history of BSE in Britain perhaps provides the most graphic example — and as virologists weigh the potential dangers posed by xenotransplantation, Hooper has delivered a timely warning.

But in pursuing his hypothesis with such zeal, Hooper risks undermining the effectiveness of this message. This week's meeting at the Royal Society in London, at which Hooper confronted those he criticizes in his book, was at times an unedifying spectacle (see page 117). There seems no prospect of a resolution of the conflict between Hooper's insistence that he has verbal evidence that chimpanzee

kidneys were sent to the Wistar Institute in Philadelphia and university research centres in Belgium, and the outright denials of the researchers involved. Similarly entrenched positions greeted the news that samples of Wistar vaccines dating from the period were produced using monkey, not chimp, tissue, and are not contaminated with the virus. The accused scientists are overoptimistic in hoping that these results will "put to rest" Hooper's accusations — the samples available today represent only some of the many batches of vaccines made. But Hooper will win few plaudits by dismissing the results as "irrelevant".

The question of whether the AIDS epidemic was triggered by contaminated polio vaccine is a legitimate one — but for reasons of public interest, not 'scientific' truth. Hooper's hypothesis cannot be tested experimentally. Given this, any inquiry into the hypothesis needs to be assessed in terms of its own social costs and benefits. If the costs — for example, in undermining public confidence in vaccines, or in distracting AIDS researchers from more pressing tasks — are too high, and the benefits questionable, it may be time for a truce. ■





Collision course
Physicists get excited
over possible sighting
of Higgs boson
p118



Off the record
Lack of data hampers
research on Gulf War
syndrome
p121



Under threat
Biodiversity faces
bleak future from
climate change
p121



In the air
First hands-on
experiment reaches
space station
p122

Tests fail to support claims for origin of AIDS in polio vaccine

David Dickson, London

The hypothesis that AIDS has its origins in a contaminated polio vaccine used in Africa in the late 1950s suffered a significant knock this week. Tests on samples of the vaccine, in storage for over 40 years, have shown no trace of HIV or its primate antecedent SIV.

The tests were conducted at three different laboratories in the United States, Germany and France (see below). The work was instigated by the Wistar Institute in Philadelphia, which has been at the centre of the vaccine controversy since it was first raised almost a decade ago.

According to the eagerly awaited results, which were revealed at a two-day meeting on the origins of AIDS held at the Royal Society in London, none of the laboratories found any trace of HIV or SIV in their samples.

In addition, studies of mitochondrial DNA from the samples failed to provide any evidence to support allegations that the polio vaccine had been prepared using chimpanzee tissue — known to be the source of the HIV virus. In each case, the material used to produce the polio virus for the vaccine was shown to be of monkey origin.

The Wistar Institute claims that the results not only vindicate its own role in the issue, but will also soothe public concern



Best of enemies: Hooper (left) claims that chimpanzee kidneys may have been used secretly to make an experimental polio vaccine in the 1950s. Plotkin (right) argues that there is “not a shred of evidence” to support such an allegation.

over the safety of vaccines, which had been raised by the allegations.

“We trust that these results will put to rest any remaining concerns of a link between a Wistar-produced oral polio vaccine and AIDS,” Clayton Buck, its acting director, said in a statement. “The findings should also serve to restore public confidence in the production and administration of vaccines.”

But writer Edward Hooper, the principal proponent of the ‘contaminated vaccine’ theory, told the Royal Society meeting that the findings made little impact on his commitment to the theory, expounded last year in his widely publicized book *The River*.

“I applaud Wistar’s decision to release batches of vaccine for testing,” Hooper said. “But different batches of vaccine were often produced in different laboratories.” Vaccine samples released did not include any from batches prepared for use in Africa, he said.

Although this week’s meeting gave detailed attention to all aspects of the origins of AIDS — including the opposing ‘cut-hunter’ thesis that the virus may have passed from chimpanzees to humans through hunting and perhaps eating contaminated animals — it was prompted largely by the public controversy surrounding the contaminated vaccine theory (see *Nature* 404, 9; 2000).

Hooper told the meeting that in his own visits to Africa over the past eight weeks, he had come up with two “smoking guns”. Both were individuals who claimed they knew that chimpanzee kidneys were being sent to

researchers in the United States and Belgium in the 1950s.

But the allegations were angrily denied by Stanley Plotkin, who had been a researcher at the Wistar Institute at the time, and had been actively involved in the African trials of the ‘CHAT’ polio vaccine.

Plotkin said that he and the former head of the Wistar, Hilary Koprowski, have questioned members of the technical staff who worked in the US laboratory at the time, and all denied having seen or heard of chimpanzee cells being used. “There is not a shred of evidence that chimpanzee kidneys were used to prepared CHAT vaccines,” said Plotkin.

Koprowski himself was similarly forthright in dismissing Hooper’s allegations and denying that he had ever used chimpanzee kidney cells. He also warned that the allegations were undermining polio vaccination efforts in Africa, pointing out that the Catholic Church in Kenya was advising parents not to have their children vaccinated because of the dangers of contamination.

Hooper denied that, in pursuing the ‘contaminated vaccine’ hypothesis, he was seeking to make anyone a scapegoat for the AIDS epidemic. But he added “we should be trying to find the truth and learn from the truth”.

Plotkin represented the issue differently. “The issue is not whether contamination with HIV might have happened, but whether it did happen,” he said. “The energy used up by this controversy should be channelled into seeking an end to the AIDS epidemic.” ■

Weighing the evidence

The tests on samples of polio vaccine produced by the Wistar Institute in Philadelphia in the 1950s (see above) were carried out at three independent laboratories, which also tested additional sets of samples and controls prepared by the Centers for Disease Control and Prevention in Atlanta, Georgia. A research group at Roche Molecular Systems in Pleasanton, California, headed by Shirley Kwok, tested the samples for the presence of SIV/HIV, and a second, headed by Svante Pääbo at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, tested for the presence and species origin of mitochondrial DNA. A third laboratory headed by Simon Wain-Hobson of the Pasteur Institute in Paris, France — one of the organizers of this week’s meeting at the Royal Society in London — performed both sets of tests.

MARTIN SHALLCROSS

Physicists haunted by hints of Higgs boson

Alison Abbott

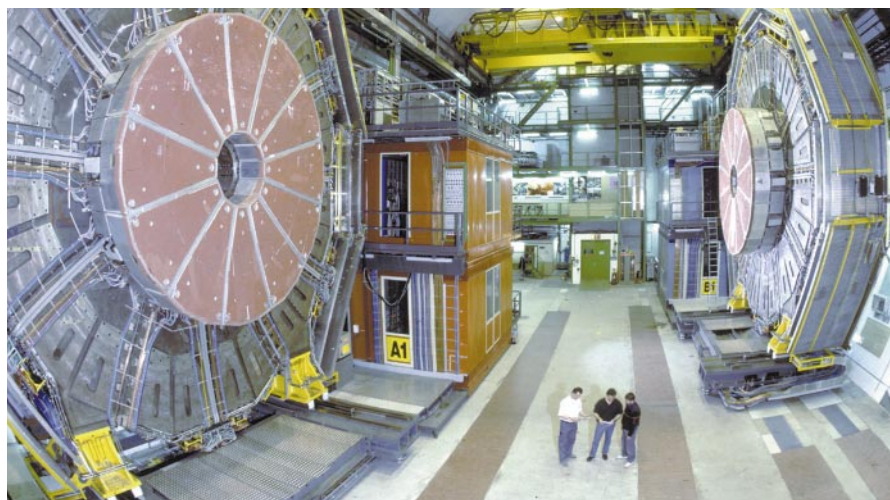
Weeks from being consigned to history, the Large Electron-Positron collider (LEP) in Geneva has put physicists in a fluster. Suggestions that the LEP has detected the Higgs boson may this week win the collider a one-month reprieve.

On 14 September, Luciano Maiani, director-general of CERN, the European Laboratory for Particle Physics near Geneva, must decide whether to switch off the LEP at the end of this month as planned. Pulling the plug would deprive CERN of its chance to follow up the intriguing hints that the Higgs particle has indeed been sighted (see *Nature* 407, 8; 2000).

According to theory, the Higgs boson interacts with subatomic particles to give them mass: the stronger a particle's interaction with the Higgs, the greater its mass. Theorists have predicted that the Higgs itself has a mass of between 50 and 150 giga-electron volts ($1 \text{ GeV} = 10^9 \text{ eV}$), with the greatest probability that the true figure lies somewhere in the middle of this range. By tuning up the LEP's operating energy, CERN scientists earlier this year narrowed the search considerably by showing that the Higgs definitely does not have a mass below 106 GeV.

The LEP's successor, the Large Hadron Collider, will run at such extremely high energy levels — 14 TeV ($14 \times 10^{12} \text{ eV}$) — that it is certain to detect the Higgs, if it exists, even if the particle's mass is as high as 150 GeV. CERN is committed to bringing the LHC online by 2005, and construction is due to begin as soon as LEP is switched off.

But the recent results from two of the LEP's four detectors have now thrown these plans into turmoil. In the past few months, the LEP has been tweaked to run at higher



Spotted: the huge detectors of the ALEPH experiment have found possible signs of the Higgs.

energies than ever. "We were not surprised to finally see hints of Higgs at a mass around 115 GeV, since this is what theory predicts," says Patrick Janot, LEP physics coordinator. "What surprised everyone was LEP's capability of working at energy levels so much higher than it was originally designed to do: it is a miracle." But for CERN's management, that miracle presents a dilemma, as it is loath to sanction any delays to the LHC.

What the CERN scientists believe they have observed is not the Higgs itself, but its decay products. When the LEP forces an electron and a positron to collide, the resulting energy is converted into matter. Occasionally, this matter will be the Higgs boson, always partnered by the Z particle. This pair immediately decays to other matter and anti-matter particle pairs. According to theory, there is an 85% chance that the Higgs decays into bottom (b) quarks. But Z particles are

produced routinely by the LEP and have numerous decay modes. The problem is that 15% of Z decays also result in b quarks, "thus faking Higgs bosons", says Janot.

High-energy physicists are excited about three b-quark events recorded from the ALEPH experiment and two by the DELPHI experiment, which stand out from the background of Z-particle decays. The experimental teams calculate that there is a 99.5% chance that the events represent new physics, with observation of the Higgs being the most likely candidate. But for physicists to accept unambiguously that the Higgs has been spotted, the chances of there being another explanation would have to be reduced to one in ten million.

If the LEP were to be run until the end of October, and no more credible events were recorded, then the idea that the Higgs has a mass of 115 GeV would be scotched. If the intriguing events continue to turn up at a similar rate, the chances that there is an explanation other than the Higgs would decrease to a few in 100,000. Janot suggests that there might then be a case for extending the LEP into next year. A one-month extension is quite possible, but sources close to CERN's management dismiss the possibility of any further delays, as that would seriously derail plans for the LHC.

If the LEP did produce a strong suggestion that the Higgs has a mass of around 115 GeV, it would allow physicists working at the Tevatron accelerator at Fermilab, near Chicago, to focus their efforts on the relevant energies. Tevatron is not optimized to searching for the Higgs. "For us, it's more like looking for Higgs in a garbage dump," says Hendrik Weerts of Michigan State University, co-spokesman for the Tevatron's D_0 experiment. Even so, if the CERN findings are real, he believes the Tevatron could pin down the Higgs within three years — well before the LHC is finished.

Gore pledges support for bioscience

Meredith Wadman, Washington.

Democratic presidential candidate Al Gore has proposed that, if elected, he will double cancer-research spending over five years and create a new bioengineering institute at the National Institutes of Health (NIH).

An economic plan released on 6 September also calls for an increased emphasis on bioinformatics. It proposes the creation of 20 university-based centres of excellence in biomedical computing.

Gore's plan calls for an \$18 billion, ten-year increase in spending on "cancer and other medical research" as part of "a larger commitment to all medical research". The National Cancer Institute's current budget is \$3 billion.

The plan does not specify the size of the larger commitment. But Devona Dolliole, deputy national spokeswoman for the Gore campaign, says that Gore plans to double NIH spending on extramural research — an increase of \$83.9 billion — although she could not say over what time frame.

The 191-page plan says that, along with the cancer money, Gore would provide "similar increases in funding for biomedical research to fight other diseases — from Alzheimer's and Parkinson's to diabetes and HIV/AIDS". His proposal says that a new bioengineering institute at NIH could lead to breakthroughs such as a cure for spinal-cord injuries, and production of artificial retinas and kidneys.

Heavyweight academies take up the cause of US postdocs

Jessa Netting, Washington

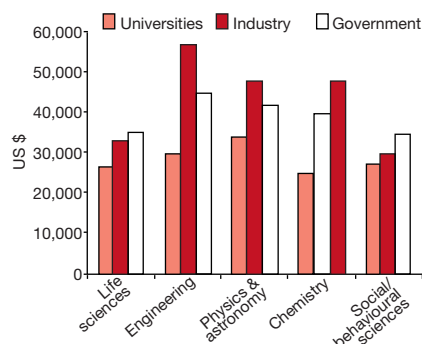
Beleaguered postdoctoral researchers in the United States have acquired some influential allies. This week, the nation's academies of science, engineering and medicine released guidelines intended to improve the lot of postdocs — many of whom are frustrated by low pay, long hours and inadequate recognition for the work they do.

Postdoc groups hope this high-level endorsement of their complaints will give them leverage in negotiating for improved conditions. "I think it will make a fantastic difference," says Pauline Wong, a postdoc in biological chemistry at Johns Hopkins University in Baltimore and president of its postdoctoral association. "It will open the eyes of institutions for whom postdocs' concerns are not on the radar screen."

Postdocs exist in a netherworld between graduate students and faculty members, usually with no official status. As a result, they often miss out on the benefits assigned to these groups. Graduate students, for instance, receive training and career guidance, whereas faculty benefit from health insurance. And both students and faculty members can turn to established grievance procedures if problems arise. But postdocs say that their uncertain status leaves them vulnerable to neglect and exploitation by unscrupulous lab chiefs.

Responding to widespread complaints — and a few horror stories — the National Academy of Sciences, the National Academy of Engineering and the Institute of Medicine asked their Committee on Science, Engineering, and Public Policy to investigate the issue. Over the past year, the committee has conducted focus groups, issued questionnaires and studied earlier surveys in compiling its report, *Enhancing the Postdoctoral Experience for Scientists and Engineers*.

The report lists ten 'action points' for urgent attention (see box). It stresses that



Academic postdocs' pay lags behind other sectors.

postdoc positions should be apprenticeships for a research career, not simply support for individual projects, and argues that postdocs should receive appropriate financial compensation. But it stops short of recommending a minimum salary.

But will employers act on the recommendations? Kimberly Paul, a biochemistry postdoc and vice-president of the Johns Hopkins Postdoctoral Association, believes the guidelines would have real teeth if grant-giving agencies wrote them into their conditions for receiving funding. The report encourages the National Institutes of Health (NIH) and the National Science Foundation (NSF) to do this.

The NIH is changing the regulations for postdocs at its campus in Bethesda, Maryland. In fact, says Michael Gottesman of the NIH Office of Intramural Research, virtually all the recommendations are already in force for the agency's own postdoctoral employees.

But neither the NIH nor the NSF have done anything to enforce guidelines for postdocs employed on grants they award to universities. "This is not an enforcement agency," says William Noxon of the NSF's Office of Legislative and Public Affairs. "We can't really do anything."

► <http://national-academies.org/postdocs>

France finally picks Parisian site for new synchrotron

Declan Butler, Paris

France this week decided to construct the proposed third generation synchrotron, Soleil — a decision a decade in the making. Paris beat Lille in the bid to host the FFfr1.65 billion (US\$218 million) facility. Construction will begin in autumn 2001 and be completed in 2005.

Never in France has a decision over a large science facility been so politically controversial. Three years ago it seemed that Soleil was on the brink of approval from prime minister Alain Juppé's government, and would be sited in Bordeaux, where he was mayor.

Then the government was beaten by the Socialists in a snap election. Last August, research minister Claude Allègre abandoned Soleil, deciding instead to contribute to the planned synchrotron Diamond outside Oxford, England — provoking strong protests from French scientists. But Roger-Gérard Schwarzenberg reopened the case when he took over from Allègre last April.

Speaking at a press conference in Paris on Monday, Schwarzenberg said that he was convinced that Soleil was "scientifically necessary" and "financially possible", emphasizing that, unlike other large facilities, synchrotrons serve many disciplines. Prime minister Lionel Jospin has endorsed the decision.

No fewer than 11 regions bid to host the machine, with three, including Paris, each putting cheques of FFfr1.2 billion on the table. Lille said it would pay whatever it was asked to.

The contest was close until the end, and many scientists felt that the northern city of Lille would win, despite the scientific arguments for building Soleil on the Saclay Plateau near Paris — home of the LURE synchrotrons which it will replace, along with many high-technology companies and major research centres.

Lille's case was strengthened by France's keenness to decentralize science away from the Paris region. The city also had an ally in Martine Aubry, who is scheduled to quit her post as minister of social affairs and become the Socialist party's candidate for mayor of Lille.

Schwarzenberg defended the choice of Saclay by referring to the three main criteria used by the working group that advised the government on Soleil. These are the geological and general quality of the site, accessibility, and the quality of the scientific environment. Unlike

Actions to improve the postdoc's lot

- Award institutional recognition, status and appropriate financial compensation.
- Develop distinct policies and standards for postdocs.
- Develop mechanisms for regular communication with advisers, institutions, funding organizations and disciplinary societies.
- Monitor and provide formal performance evaluations.
- Ensure postdocs have access to health insurance.
- Set limits for total time of postdoc appointment.
- Invite participation of postdocs when creating standards, definitions and conditions.
- Provide substantive career guidance.
- Improve quality of data on postdoctoral working conditions and employment prospects.
- Improve the transition of postdocs to regular employment.



Aubry: lobbied in vain for Lille.

► Allègre, who revealed few details regarding his decisions on Soleil, Schwarzenberg made the report public.

Taking a swipe at his predecessor, who announced Soleil's abandonment during August, when France is on holiday, Schwarzenberg said he had waited until September to "to avoid any announcement in the month of August".

The regions will provide an unprecedented proportion of Soleil's costs; the state will end up paying just FF200 million–300 million. The regions already contribute towards the cost of the planned national university, and Schwarzenberg hinted that regional money would play an increasing role in the research budget.

The machine will have an energy range of 2.5 to 2.75 GeV, producing beams from the ultraviolet to both soft and hard X-rays. Four beamlines will be dedicated to biological research instead of the two proposed originally. Twenty-four beamlines will be built at first, increasing to 40.

Soleil will have a company's legal structure, enabling it to function continually and offer higher salaries: rules at public research organizations would not allow such flexibility. ■

Drug-resistant HIV shows a worrying increase in the UK

Natasha Loder, London

The United Kingdom has seen a large increase in new cases of multidrug-resistant HIV over the past five years, according to a researcher at the Public Health Laboratory Service (PHLS) who is calling for wider testing for resistance.

Multidrug-resistant HIV is resistant to one or more members of all three classes of anti-HIV drug. It develops in patients who are receiving therapy, and is transmitted by normal HIV infection routes. This is the first time such an increase has been seen.

Deenan Pillay, of the PHLS's Antiviral Susceptibility Reference Unit at the University of Birmingham, presented the figures at last week's British Association science festival at Imperial College in London. The data are from a register of patients who had recently contracted HIV, all testing HIV-negative in the preceding 18 months or showing recent seroconversion at the time.

Pillay found that five out of 24 patients who had been infected up to May 2000 had key resistance mutations. In the previous five years, only one in 40 was similarly infected.

The news will disappoint many scientists and clinicians. High-level resistance mutations are less prevalent in studies in other countries, leading some to hope they would remain rare. Robert Shafer, an assistant professor of medicine at the University of Stanford, California, is cautious about interpreting unpublished figures.

But he says that "if correct, they are important and distressing," because they suggest that it is not just those unaware of their infection who are spreading HIV, but also people who know of their condition and have access to treatment.

Resistance often develops because



patients cannot maintain the harsh drug-taking regime. But even those with full compliance have a 20% failure rate in the United Kingdom, possibly because the treatment is unable to halt viral replication entirely.

Pillay told the meeting that "the emergence of drug resistance is virtually inevitable" and warned that the spread of resistant strains in the United Kingdom threatened "the success we have seen in anti-viral therapy to date". New classes of drug could offer hope to patients with multidrug-resistant HIV.

There are currently 14 anti-HIV drugs in three classes: nucleoside analogue reverse-transcriptase inhibitors, non-nucleoside reverse-transcriptase inhibitors, and protease inhibitors. Despite rising numbers of HIV cases, the number of deaths from AIDS in England and Wales has halved since 1995 — partly owing to the introduction of protease inhibitors in 1996 as part of combination therapies that suppress viral replication.

Once a virus develops resistance this often confers high- or low-level resistance to other drugs in the same class. But occasionally a mutation can make the virus particularly sensitive to another drug in the same class.

Shafer has created a database linking data on the sequences of HIV-1 reverse transcriptase and protease, drug treatment histories and drug susceptibility, which he hopes will help in the design of treatment regimes. Genetic tests can show why patients are failing therapy and could help doctors design the best long-term drug treatment regime.

Pillay, whose unit conducts such tests, says this "is now recommended in order to guide the most effective therapies". But each test costs about £300 (US\$426), and some National Health Service trusts are refusing to pay for them. Shafer says that preliminary data suggest that the cost of testing are offset by a decreased use of ineffective drugs. ■

Neuroscientist to lead Salk Institute

Rex Dalton, San Diego

Ending a year-long search, neuroscientist Richard Murphy has been named the new president and chief executive of the Salk Institute in La Jolla, California. Currently director of the Montreal Neurological Institute at McGill University in Canada, Murphy will take up the post on 1 October.

Born in Massachusetts, Murphy earned his doctorate in zoology from Rutgers University, and worked at Harvard University and the University of Alberta in Edmonton before moving to McGill in 1992. His main research has been into the way

neurotrophins — proteins that promote the growth and survival of nerve cells — are produced in the brain.

"He has shown considerable skill for attracting new resources and talent and generally raising the level of facilities and research wherever he has been," Frederick Rentschler, chairman of Salk's board, said in a statement.

Murphy will have to keep established scientists happy while encouraging new talent. Internal conflicts drove previous president Tom Pollard to give up the post last year (see *Nature* 400, 805; 1999). ■

US panel draws blank on Gulf War symptoms

Meredith Wadman, Washington

The fierce controversy over the origins of the chronic symptoms reported by US Gulf War veterans is set to continue. Last week, an expert committee at the Institute of Medicine (IOM) concluded it could not resolve the issue because there is not enough published evidence about the effects of exposure to suspect substances.

The committee based its opinion on a close reading of 1,000 papers from the scientific literature. These were almost all reports of occupational, clinical and terrorist exposure, rather than of Persian Gulf veteran exposure.

It concluded that the papers do not provide sufficient data to place blame with any certainty on the substances of most concern to veterans — the nerve gas sarin; pyridostigmine bromide (PB), a medicine troops took prophylactically to blunt the effects of nerve gas; depleted uranium from tanks and munitions destroyed in the war; and vaccines given to some troops to prevent anthrax and botulism.

"The bottom line is that we simply don't know enough to say whether there is a connection between exposure to these agents, or combinations of these agents, and specific health outcomes that remain long after the exposure," Harold Sox, who chaired the committee, told a news conference in Washington. Sox is professor of medicine at the Dartmouth-Hitchcock Medical Center in Lebanon, New Hampshire.

The committee said that a lack of information about actual troop exposures meant it had to review papers that did not involve Persian Gulf veterans. Even in this literature,



Chemical conundrum: too few data are available to establish the long-term effects of nerve gases.

insufficient information about the chronic effects of exposure made it impossible to draw conclusions in most instances.

There were some exceptions. The committee has now ruled out any role for depleted uranium in causing kidney disease or, in low doses, lung cancer. Conversely, three papers showed that neurological and psychological symptoms could persist for at least six months after exposure to sarin.

The committee also challenged several studies of Gulf War veterans by Robert Haley, director of epidemiology at the University of Texas Southwestern Medical Center in Dallas (see *Nature* 385, 187; 1997). These suggested

that chronic neurological damage might be related to PB. But the IOM study noted methodological flaws in the studies, including heavy reliance on self-reporting, a small survey sample, and lack of a control group.

The IOM's report drew immediate criticism from activist veterans. They suggested that proof of links between their wartime exposure and current symptoms might lie in classified US Department of Defense information to which the committee did not have access. They also criticized the fact that the report did not consider exposure to combinations of agents, which they argued was a more likely battlefield scenario. ■

Global warming threatens extinction for many species

David Spurgeon, Montreal

Rapid rates of global warming will be bad news for biodiversity, particularly in the far north, according to a report prepared for the World Wildlife Fund and released recently in Toronto.

Jay Malcolm, of the University of Toronto's forestry faculty, and Adam Markham, of the non-governmental organization Clean Air-Cool Planet, based in Portsmouth, New Hampshire, warn that many species may be unable to migrate "away from increasingly less favourable climatic conditions to new areas that meet their physical, biological, and climatic needs".

The researchers used seven climate models and two vegetation models to analyse how fast species might have to move to keep up with projected warming.

High migration rates will be required

over 38.3% of the land surface of Russia and 33.1% of Canada. The authors conclude that species living in the tundra and in coniferous forests "may be among the most vulnerable to global change".

In many countries, including Sweden, Finland and Iceland and several former Soviet republics, more than half of the existing habitat could be lost or changed into another habitat type. The same goes for seven of the twelve Canadian provinces. And in the United States, more than a third of existing habitat in eleven states could change.

Noting that even relatively optimistic predictions suggest that carbon dioxide concentrations are likely to have doubled by around the middle of this century and almost tripled by 2100, the researchers suggest that species may need to move even faster than their findings indicate.

"If past fastest rates of migration are a good proxy for what can be attained in a warming world, then radical reductions in greenhouse gas emissions are urgently required in order to reduce the threat of biodiversity loss," the authors conclude. ■



Changing homes: tundra-living species such as these caribou are threatened by climate change.

NASA encounters biggest-ever Antarctic ozone hole

Washington The largest-ever Antarctic ozone 'hole' has been detected by the US space agency NASA. It measures 28.3 million square kilometres — or three times the size of the United States. The Antarctic ozone hole opens every year around September and October. The previous record size was seen two years ago, when the hole extended over 27.2 million square kilometres. This year's hole could continue to increase in size over the next month.

Ozone-destroying gases persist in the atmosphere. So although international agreements under the Montreal Protocol have curtailed their production, concentrations in the stratosphere are only now beginning to peak and it could be many decades before the hole is no longer an annual occurrence.

Study suggests that France faces its own BSE epidemic

Paris Preliminary results of a French screening programme suggest that up to 1.5 in every 1,000 cattle in the country are contaminated with BSE. The agriculture ministry has tested 6,000 cattle in western France. Only those not destined for the table — that is, cattle found dead or sick — were tested; eight showed signs of BSE infection. Final results of the programme, which will extend to some 40,000 animals, will be issued in six months.

The rate of infection appears similar to that revealed by an earlier investigation in Switzerland. This also looked at animals entering slaughterhouses. The French newspaper *Le Figaro* estimates that 1,200 contaminated animals are consumed in France each year, despite attempts to curb the epidemic. The ministry disputes this figure, however.

European Parliament condemns cloning

Strasbourg The European Parliament last week urged members of the British Parliament to reject a proposal by their government that they should revise legislation on embryo research to allow research into 'therapeutic cloning' using cell nuclear transfer (see *Nature* 406, 815; 2000).

After a debate in which one French member of the parliament had described the British proposal as "monstrous", and an Italian colleague as an "affront to civilization", the parliament approved by 237 votes to 230 a resolution which also demanded that the cloning of human embryos "at all stages of their development"

be outlawed throughout Europe. It also voted that such a ban should be included in the European Union's Charter of Fundamental Rights.

Space station gears up for a life in science

Washington The first hands-on experiment to be conducted aboard the International Space Station has been launched by the space shuttle Atlantis, which docked with the station on 10 September.

Developed by scientists working at the



Good vibrations: MIT researchers display their apparatus to study movement in zero gravity.

Massachusetts Institute of Technology and the Air Force Research Laboratory in Albuquerque, New Mexico, the Middeck Active Control Experiment II (MACE II) will study the effects of vibrations in zero-gravity conditions. The goal is to improve the design of telescopes, robotic arms and other devices.

MACE II is expected to start in early November — when the first astronauts are due to arrive on the station. Bill Shepherd, the commander of the first crew, will have primary responsibility for running the experiment. Another experiment, on crystallized plasmas, is due to start at about the same time (see *Nature* 405, 7; 2000), but will not require such active astronaut involvement.

Pay boost for UK medical researchers

London The UK Medical Research Council (MRC) has negotiated a large one-off pay supplement for its non-clinical researchers. Some 1,000 scientists are being offered a 9% supplement, to be staged over three years from 1 August 2000. The increase will come from the MRC's existing budget, but had to be approved by the government.

In addition, all council staff are being offered performance-related salary increases starting at 3.8% — well above the current rate of inflation. Union members will vote on the settlement on 24 September.

"We are very pleased," says George Radda, the MRC's chief executive. "I see this as recognition by the government that scientists in this country need to be

rewarded for their role." He hopes universities will now follow the MRC's lead.

Genome summit will plot best path to the finish

Washington The leaders of the international effort to sequence the human genome are to meet in the Paris suburb of Evry tomorrow (15 September) to plot their strategy for 'finishing' the genome. Currently, only 25% of the sequence is in an accurate, assembled form.

Francis Collins, who as director of the US National Human Genome Research Institute heads the project, says that the scientists will try to ensure that no part of the sequence lags behind and that none are duplicated. The 'G-6' sequencing consortium is led by the United States and Britain, and includes Germany, France, Japan and China.

Top Spanish science job goes to physicist

Barcelona Physicist Rolf Tarrach was last week named as the new president of Spain's Higher Council of Scientific Research (CSIC). The CSIC is the country's main research organization, and employs about 2,000 scientists in more than 100 institutes throughout the country. Tarrach, who is professor of physics at the University of Barcelona, will take over from the present director, César Nombela.

Tarrach is a vociferous critic of the way that universities operate and their lack of sufficient research funds. After his appointment was announced, he said that a high priority for CSIC should be to increase the recruitment of young postdoctoral researchers who had received training abroad. He also promised to promote closer collaboration between researchers in CSIC and the universities.

US nuclear workers lacked protection and information

Washington A study released in Washington last week argues that the US government failed to give nuclear-weapons workers in the 1940s and 1950s adequate protection from high levels of radiation. It also says that managers at privately owned nuclear weapons plants deliberately misled employees about the dangers of their working conditions.

The study was carried out by the non-profit Institute for Energy and Environmental Research, and financed by the newspaper *USA Today*. It found that, contrary to assurances to workers that they did not face an unusual hazard, radiation levels within the plants routinely reached 500 times allowable limits. The institute is urging the US government to investigate the problem further, and provide free medical care to ageing former nuclear workers.



FRANK SPOONER PICTURES

What price the Olympian ideal?

Without more money for their research, antidoping scientists will continue to be beaten into second place by pharmaceutically assisted athletes. Alison Abbott reports from the front line of sport's drugs war.

Cheating in sport is nothing new — only its form has changed. In ancient Greece, athletes attempted to lay their rivals low with curses, or simply turned to bribery. Fines levied on those caught were used to build bronze statues of Zeus, which lined the entrance of the Olympic stadium, reminding competitors of the perils of cheating.

Today, cheats rely on chemical assistance, using an array of banned substances to get more from their bodies than nature and training would otherwise provide. And although those who are caught face lengthy bans, the depressing truth is that pharmaceutically enhanced athletes are usually one step ahead of the scientists employed to catch them. If the International Olympic Committee (IOC) and other sports governing bodies do not provide more funding to back their tough words on drugs, these scientists say they will continue to be runners-up in the doping contest. "Drug testing costs a lot of money and sports federations are never keen payers," says Mats Garle, scientific director of the Swedish Doping Control Laboratory in Huddinge, near Stockholm.

The latest weapon in the fight against

doping comes from Australian and French researchers who have developed methods for detecting abuse of erythropoietin (EPO), a hormone that boosts the body's production of red blood cells. Approved by the IOC in late August with great fanfare, these two tests will be used at the Sydney Olympics, which open on 15 September. But few athletes are likely to be caught out at the games. Provided they stop taking EPO a few days before their events, they will probably test negative. In fact, the tests are likely to be most useful in out-of-competition testing, which has already begun — but it is rumoured that the EPO abusers are already working on ways to dodge detection.

Meanwhile, progress towards tests for another performance-enhancing drug, human growth hormone (hGH), has stalled. Although the scientists developing the tests are confident that they could have been readied in time for Sydney, the IOC refused to sanction funding for the necessary validation studies. As a result, unscrupulous athletes know that they can abuse hGH with impunity.

Before any new drugs test is introduced, sports governing bodies must be convinced

that it will stand up to legal challenges. If the doping control authorities lose a courtroom battle, the consequences can be disastrous. In 1997, the British Athletics Federation went



Back on track: sprinter Merlene Ottey successfully overturned a drugs-related ban.

MARTIN MEISSNER/AP

In a spin: the 1998 Tour de France was reduced to chaos (left) after drugs were found in the possession of Willy Voet, a Festina team employee.

bankrupt, partly as a result of court costs incurred after the middle-distance runner Diane Modahl challenged its decision to ban her following a positive test for testosterone. Modahl convinced the court that bacterial growth caused by a failure to refrigerate her urine sample properly could have led to a false positive result.

Hormone hunting

Given this, the IOC's decision to adopt the two EPO tests for the Sydney games is a genuine milestone in antidoping science. Recombinant EPO, a peptide of 165 amino acids produced by genetic engineering, is among the world's top selling pharmaceuticals. It is used to treat anaemias associated with disorders such as kidney failure. By boosting the production of red blood cells, it increases the flow of oxygen to the muscles. Some competitors in endurance sports such as cycling and distance running probably started using EPO as soon as it became available in 1987. Until now, there has been no way to detect this abuse.

In approving one of the EPO tests, the IOC is for the first time requiring athletes to give blood samples for doping control. Developed by Michael Ashenden and his colleagues at the Australian Institute of Sport near Canberra, the test measures the EPO concentration in blood as well as four other factors affected by raised EPO levels¹. Precursors of red blood cells, known as macrocytes and reticulocytes, are overproduced in bone marrow when EPO levels are raised, and they leak out into the circulation. So Ashenden's test measures the levels of red blood cells and these two precursors. It also measures the serum concentration of a protein called soluble transferrin receptor, which is involved in iron metabolism — and as such influences the production of the oxygen-carrying haemoglobin complexes found in red blood cells.

The other test, described earlier this year in *Nature*² by Françoise Lasne and Jacques de Ceaurriz of the French National Anti-Doping Laboratory in Châtenay-Malabry, near Paris, detects directly the presence of recombinant EPO in urine. The test is based on a subtle difference between human EPO and that produced *in vitro* for pharmaceuticals. The recombinant EPO has the same amino-acid sequence as the natural hormone but, because it is produced from non-human cells, it has a different number of sugar residues attached to it. As a result, the electrical charges on the two forms of EPO are different and they can be separated using an electrophoretic technique called isoelectric focusing.

To avoid the possibility of false positive

The lot of the dope police

Unlike the lives of the athletes they police, the work of the antidoping scientists is far from glamorous. Thousands of samples, each held in tamper-proof bottles, must be subjected to a set of standardized analyses. It is easy to get bored with the routine, says Wilhelm Schänzer, who heads the International Olympic Committee (IOC)-accredited doping control lab in Cologne. Last year, the 35 scientists in his lab tested 10,800 samples, 159 of which proved positive. Two-thirds of the positives were for anabolic steroids, most of the rest for stimulants such as ephedrine.

Each sample is divided into two, so that a retest can be done in the event of a positive result. For the retest, the athlete involved, or witnesses representing them, can attend. This is always "uncomfortable" for the scientists, says Jordi Segura, head of the IOC-accredited lab in Barcelona.

Doping control labs are also always on the alert for biochemical clues that athletes may have started to abuse a new compound for which no test is available, or which may not yet have been put on the banned list. Suspicions are usually aroused by the presence of unknown nitrogen-containing molecules in an athlete's urine. In some cases, it is possible to predict which drugs they are likely to experiment with. "We do try to anticipate the abuse potential of new drugs which come onto the market," says Schänzer.



Testing times: Schänzer says that doping control can be a frustrating task.

What the antidoping scientists find most frustrating is the validity of their work being challenged by athletes' lawyers. "A new approach or methodology may take scientists years to develop, yet as soon as there is a positive test all the science is thrown into question," says Francesco Botrè, head of the IOC-accredited doping control laboratory in Rome. And if a disputed positive test ends up before an international arbitration panel, or even in court, it can be a harrowing experience for the scientists on both sides. "It is not that easy being beaten for hours by a barrister in a court-like setting," says John Honour, an endocrinologist at University College London who on several occasions has acted as an expert witness for accused athletes.

results in Sydney, sanctions will only be taken against athletes who fail both tests. The blood changes tested for by Ashenden's method linger for two to three weeks after an athlete stops taking EPO, but the recombinant EPO itself is flushed out of the body within a few days. De Ceaurriz suspects that the performance-enhancing effects of EPO start to diminish three days after its last administration. Even so, he believes athletes can gain an advantage if they use EPO up to a few days before their competitions, and yet still test negative for the urine test. He says the tests will be most valuable in out-of-competition monitoring, where samples can be demanded at any time.

After watching the performance of the EPO tests in Sydney, IOC-accredited labs will decide whether to include them in their battery of standard tests. At the moment, these cost around 150 euros (US\$131) for competition samples and 100 euros for out-of-competition testing. The two EPO tests could add considerably to the costs. Wilhelm Schänzer, director of the IOC-accredited doping control laboratory in Cologne, calculates that Ashenden's blood test alone would add around 25 euros, not including the costs involved in blood collection.

The intensity of the war on doping will always depend on money, both for carrying

out the tests and for developing new ones. In 1998, the IOC decided to fast-track the development of the EPO tests by investing US\$1 million. This figure was matched by the Australian government.

A growing problem

But the IOC's antidoping research budget is so limited that this decision meant it had to abandon promising research into tests for hGH. Used to treat dwarfism, recombinant hGH first made its appearance in the mid-1980s. As usual, athletes recognized its fat-burning and anabolic effects long before the scientists. Even before the recombinant hormone became available, a publication called *The Underground Steroid Handbook* noted that hGH was already established in power-lifting and "within a few years will be a commonly used drug in all strength athletics". This prediction seems to have come true. The drug is often stolen from manufacturers and hospital pharmacies, and athletes have been caught carrying supplies. In February this year, 1,575 vials of hGH were stolen from a pharmaceutical importer in Sydney.

In the mid-1990s, the IOC and the European Commission co-funded a three-year international project called GH2000 to develop tests to detect this substance. The

A storm over steroids

Until last year, few people — apart from scientific specialists and those associated with top-level sport — had heard of nandrolone. But in the run-up to the Sydney games, this anabolic steroid has repeatedly hit the headlines following the revelation that some of track and field's biggest names have tested positive for the drug — among them Linford Christie, 100-metres sprint champion at the 1992 Barcelona Olympics (pictured).

The resulting controversy has seen the accused athletes and their lawyers question the work of the antidoping scientists and has set Britain's athletics federation against its international parent body. It has also drawn attention to the nutritional supplements consumed by top athletes, some of which

contain traces of nandrolone — even though it is not mentioned on the labelling.

Nandrolone is just one of many synthetic anabolic steroids used in medicine. But they can also be used by athletes, and nandrolone has long been on the International Olympic Committee's list of banned substances. The test used to identify the cheats uses mass spectrometry to detect the main metabolite of nandrolone, 19-norandrosterone, in athletes' urine.

Human urine can contain tiny quantities of naturally produced 19-norandrosterone. So the IOC-accredited scientists who devised the nandrolone test conducted validation studies to determine threshold levels of the compound above which an individual can reliably be assumed to have taken the drug or one of its precursors. These have been set at 2 nanograms per millilitre for men, and 5 ng ml⁻¹ for non-pregnant women. The current controversy centres on the validity of these thresholds, with accused athletes — and UK Athletics, which took over as the sport's governing body in Britain after the British Athletics Federation went bankrupt — arguing that individuals could record results above the thresholds without taking any banned substance.

Some athletes have appealed successfully against positive nandrolone tests. In July, the Jamaican sprinter Merlene Ottey was cleared by an International Amateur Athletic Federation (IAAF) arbitration panel after questions were raised about the way her positive test was done. In particular, it emerged that Ottey was dehydrated when she gave her sample — which would have elevated the concentration of 19-norandrosterone in her urine — and this had not been taken into account.

Appeals that seek to undermine the basis for the thresholds are much more ambitious. But that is what Christie and two other British athletes, European 200-metres champion Dougie Walker and 400-metres hurdler Gary Cadogan, tried to do last month, when their cases were heard by an IAAF arbitration panel. All three had

tested above the threshold for 19-norandrosterone: Cadogan at 10.6 ng ml⁻¹; Walker at 12.6 ng ml⁻¹; and Christie at 200 ng ml⁻¹. With the backing of UK Athletics, which had cleared them of nandrolone abuse, the athletes' arguments rested on unpublished research conducted by Ron Maughan, a physiologist at the University of Aberdeen and a nutritionist for the British Olympic team.

In a project commissioned by UK Athletics and the IAAF, Maughan studied several athletes who had tested positive for nandrolone and a group of physically active, healthy volunteers. The subjects were not monitored continuously, and Maughan does not place too much emphasis on the results from the athletes — who may have had a vested interest in the study's results. More important, he says, are the findings from the healthy volunteers who consumed large quantities of the sort of food supplements used by athletes. First, Maughan tested three volunteers and found that the one who had exercised that day tested above the threshold for the nandrolone tests. In a follow-up study, in which ten volunteers were given supplements and told to exercise, five tested above the threshold.

Maughan stresses that his results are preliminary. "This is at a very early stage and we recognize we need more information," he says. Given this, and questions surrounding the precise methods used, the IAAF panel rejected the British athletes' appeal. Wilhelm Schänzer, director of the IOC-accredited doping control lab in Cologne, which conducted Christie's positive test, argues that the nandrolone test threshold has been subject to extensive validation. "We have data from hundreds of thousands of samples," he says. These data have not been published, but Schänzer adds that a French study on 30 healthy men⁵ found a maximum urinary concentration of 19-norandrosterone of 0.32 ng ml⁻¹.

Some positive nandrolone tests could be the result of athletes buying contaminated nutritional supplements. Maughan could find no trace of nandrolone or related compounds in the supplements his volunteers consumed. But several antidoping laboratories, including Schänzer's, have analysed such supplements, procured from athletes or from suppliers advertising on the Internet, and found traces of nandrolone or its precursors not mentioned on the labelling. "We informed the IOC and the doping control officials of the different sports federations, but so far only the international cycling federation has responded by formally advising its members against using nutritional supplements," says Schänzer.

Although it might seem harsh to take sanctions against those who may have inadvertently consumed a banned substance, the rules state that it is the responsibility of athletes to ensure that they remain clean. The antidoping scientists' job, says Schänzer, "is to say whether or not certain substances are in the body, not how they might have come to get there".

► project, which concluded at the end of 1998, was led by Peter Sonksen, an endocrinologist at St Thomas' Hospital in London.

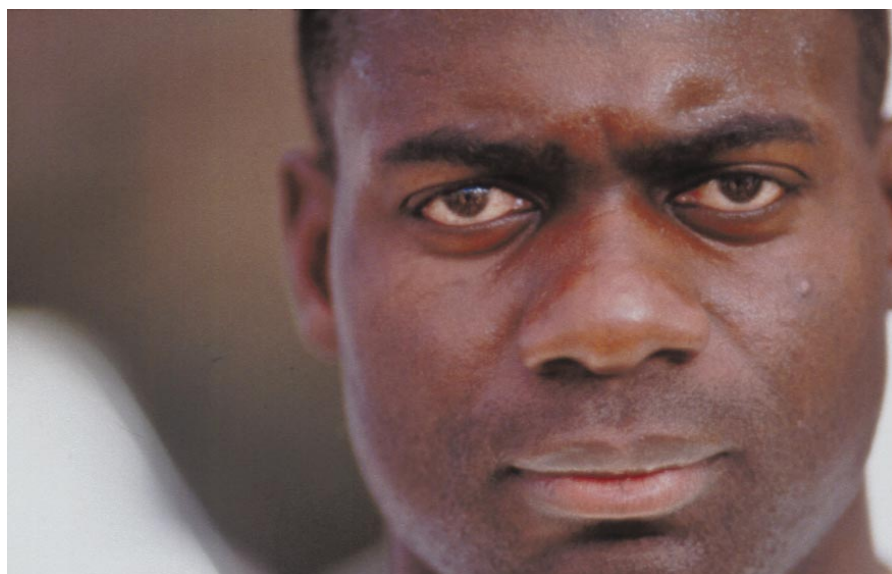
The GH2000 consortium delivered its report to the IOC in January 1999. It had developed a series of blood markers that could be used to test for elevated hGH levels, including insulin-like growth factors and proteins that bind to them³. "We have found a very credible test for exogenous hGH," says Sonksen. "It only needs to be tested out more broadly, in a range of ethnic and hormonal backgrounds, and in other conditions."

Weighing the odds

Working independently of GH2000, researchers led by Christian Strasburger at the Ludwig-Maximilian University in Munich have developed a direct test for recombinant hGH⁴. This relies on the fact that hGH exists in different molecular forms, the two major fractions of which have molecular masses of 22 kilodaltons and 20 kDa. Although only half of the body's own hGH is in the heavier form, for recombinant hGH the figure is 95%. The test uses antibodies to identify the two forms, and so allows any shift in the natural ratio to be spotted.

Sonksen costed validation studies for the GH2000 and Munich tests at around US\$5 million, and requested continued funding. The consortium also responded to a formal call for proposals for research issued by the IOC in August 1999, but was turned down. Sonksen still has thousands of samples available for analysis in his freezer. "Now GH2000 is in cold storage both literally and metaphorically," he says.

The IOC argues that Sonksen's cost estimates were too high, and had not been convincingly justified. "We won't do it at this price," says Patrick Schamasch, medical director of the IOC, based at its headquarters



Fallen hero: sprinter Ben Johnson lost his gold at the 1988 Seoul Olympics after failing a dope test.

in Lausanne, Switzerland. But Sonksen rejects this criticism. He maintains that his experience with GH2000 allowed him to give precise estimates for the development needed to bring the tests up to a court-proof standard.

The German Sports Research Institute in Cologne is now supporting further work to validate the Munich test, drawing on some of the GH2000 samples. Strasburger hopes that the newly created World Anti-Doping Agency (WADA), based in Lausanne, will follow through on statements that tackling hGH abuse will be its top priority, and provide the money needed to bring the test into general use. When it is fully established, WADA is supposed to take over doping control from the IOC. But the IOC and national sports federations will provide most of the agency's funding. And some scientists claim that if these bodies were committed to

clamping down on drug abuse, they would divert much more money into promising new detection methods. "If they were really serious, they would be well-advised to at least support the development of new scientifically validated analytical tests," says Strasburger.

That view received strong support in a report from the US National Commission on Sports and Substance Abuse, commissioned by the White House Office of National Drug Policy and released on 8 September. It criticized the IOC's "seeming lack of will" to tackle the doping problem, and called for a five-year research programme, costing up to US\$100 million, to give the antidoping labs the tools they need.

As soon as any new test is introduced, cheating athletes will seek ways to avoid detection, or find other performance-enhancing drugs for which there are no tests. Some cyclists are already believed to be experimenting with EPO precursors, and oxygen-carriers such as perfluorocarbons and oxyhaemoglobin. Drugs that increase the production of blood plasma, meanwhile, could be used to dilute out EPO before the blood tests. And if a source of recombinant EPO from a human cell-line becomes available, the urine test for EPO would immediately become ineffective.

Given the determination of their opponents, antidoping scientists realize that it will always be tough to keep pace with the drug cheats. But if they are not given a larger share of the huge sums of money circulating within the world of sport, their task will almost certainly be hopeless. ■

Alison Abbott is Nature's Senior European Correspondent.



Under siege: Schamasch (right) and Olympic medical chairman Prince Alexandre de Merode.

1. Parisotto, R. *et al. Haematologica* **85**, 564–572 (2000).
2. Lasne, L. & de Ceaurriz, J. *Nature* **405**, 635 (2000).
3. Wallace, J. D. *et al. J. Clin. Endocrinol. Metab.* **84**, 3591–3601 (1999).
4. Wu, Z., Bidlingmaier, M., Dall, R. & Strasburger, C. J. *Lancet* **353**, 895 (1999).
5. Dehennin, L., Bonnaire, Y. & Plou, P. J. *Chromatog. B* **721**, 301–307 (1999).

Step back to see how science and humanity fit in the big picture

Sir — Although your Millennium Essays focus on events in the past 1,000 years, it is easy to lose perspective when analysing time-windows much greater than our life-span. Our brains are not adapted to think in vast time-frames. So it may be helpful to take a bird's-eye view of our past millennium by using a series of time-windows — each one-hundredth the size of the previous one — to extend the analysis deeper in time.

In the largest time-window imaginable, comprising tens of millions of millennia, the past millennium doesn't show up. This window places the formation of Earth and the emergence of life in perspective: if the event creating our present Universe (the Big Bang) occurred around 15.5 million millennia ago, the development of life on Earth represents about 25% of the history of our Universe.

One per cent of this large window produces a second time-window covering the past 200,000 millennia, in which evolved extant animals and plants, primates and, finally, hominids.

In the third window, spanning 2,000 millennia, human invention starts to become apparent. Stone tools were used during the first half and fire was domesticated around the halfway point. These preceded the formation of the oldest-known fossils of *Homo erectus*, showing that creativity, technology development and invention were possessed by at least some of the australopithecine and other hominids long before *H. sapiens* appeared. Human creativity developed slowly, leaving the earliest paintings as evidence at the end of this window. This time-window saw the extinction of many other creatures (including all hominids except *H. sapiens*) — some caused by the sophisticated traps and projectiles made by *H. sapiens*.

The fourth time-window, using a 20-millennia lens, shows that both extinctions and human creativity continued to accelerate. Humans were weaving clothes at the beginning of this period, before the last ice age, by the end of which plants and animals were domesticated. This led to the development of agriculture about 10 millennia ago. The development of ceramics, the wheel, iron, bronze, steel and written documents followed at ever-shorter intervals. The use of script gives us a direct view on the thought processes of Homer, Lao-Tse, Buddha, Aristotle, Christ, Ptolemy, Muhammad and many others. These scripts show that human emotion did not differ significantly then from that experienced today.

Five per cent of this 20-millennia time-window focuses on the past millennium.

The main feature of this millennium seems to have been the emergence of the scientific method. Science, defined as the method that subordinates theory to experimental results, was pioneered by Galileo and has been widely accepted as a superior form of thought only during the past two centuries. An interesting aspect is the shift of creative power from China to the West as the nationalistic and xenophobic policies of the Ming dynasty, started by Chu Yuan-chang in 1368, hindered commerce and the expansion of creativity in China. I detect a pause in the growth of creativity in the Christian world, spanning the first millennium AD (after Ptolemy and the Roman Emperor Julianus) to the Renaissance, overlapping with bursts of creativity in the Islamic world.

The past millennium shows no sign that the exponential expansion of human mental capabilities has stabilized. Evolution of other animal species, however, suggests that exponential evolutionary development of behavioural or morphological traits eventually stops. Even human creativity may start to do so in the third millennium AD.

Klaus Jaffe

Department of Biology, Universidad Simón Bolívar, Caracas 1080, Venezuela

Patent dispute hangs over kringle 5

Sir — In June you published a News brief relating to a suit brought by Abbott Laboratories against Children's Hospital in Boston and against me and two of my younger associates¹. The suit asserted rights to a patent assigned to Children's Hospital on an endothelial and angiogenesis inhibitor called kringle 5 which was discovered in the hospital's laboratories.

Abbott's claim is limited to one small aspect of Children's Hospital work, not involving any work on angiogenesis or the discovery of the angiogenesis inhibitors angiotatin, endostatin, or others.

If an academic institution used to see its laboratory discoveries translated to the bedside, it must seek patents. Patent disputes, unfortunately, are not unusual. What is unusual here is that Abbott Laboratories, in a deliberate attempt to claim for itself a discovery that it neither made nor owns, has attacked the integrity of scientists at Children's Hospital and has publicized its false allegations worldwide.

I have no personal financial interest in this matter. My only objective is to see that medical advances are brought to clinical application so that patients can benefit. It will be reprehensible if Abbott's actions disrupt or delay this.

Abbott's claims against Children's Hospital, me and my associates are false and

malicious. The truth of the matter is as follows. First, Abbott based many of its claims in court on an alleged confidentiality disclosure agreement and attached the agreement as a key part of its court filings. The alleged agreement was originally sent to Children's Hospital in June 1995, but it was never agreed to by the hospital nor returned to Abbott. The hospital's Office of Technology rejected the sweeping agreement. A technology office official wrote many comments on the document and in July 1995 told Abbott and its employee Donald Davidson that the agreement was unacceptable. He forwarded to Abbott an alternative agreement, but Abbott never signed it.

Second, five months later I received a letter from Abbott clearly acknowledging the lack of any agreement with Children's Hospital and proposing that, in the absence of a legal agreement, we enter into a "gentlemen's agreement" whereby Abbott would provide plasminogen fragments in exchange for my laboratory continuing to provide Abbott with expertise and materials. Abbott also proposed that Davidson and another Abbott employee visit my laboratory to learn more. I informed Abbott that we did not wish their employees to visit, but that we would be happy to share our expertise and cell lines with Abbott, which we did.

Third, Davidson was working on the thrombolytic (clot lysing) properties of plasminogen and had no idea that any fragment of plasminogen had anti-endothelial or anti-angiogenic activity until after my laboratory published the discovery of angiotatin in *Cell*² in October 1994, six months after Yihai Cao in my laboratory first requested plasminogen fragments from Davidson and other scientists, including Miguel Llinas of Carnegie Mellon University and Stephen McCance of American Biogenetic Sciences and Notre Dame.

Fourth, the anti-endothelial and anti-angiogenic properties of various kringle fragments of plasminogen were discovered by Cao, Michael O'Reilly and me. Cao and O'Reilly were postdoctoral fellows in my laboratory at Children's Hospital.

Fifth, Cao published two papers on his results^{3,4}. He graciously included Davidson, McCance and Llinas as co-authors because they supplied plasminogen fragments or advice on preparing those fragments.

Last, on 18 July 2000, Children's Hospital filed a detailed answer and counter-claims, and on 14 August, Abbott filed a pleading in court admitting that it had doctored the alleged agreement it had attached to its complaint.

Judah Folkman

Children's Hospital, 300 Longwood Ave, Boston, Massachusetts 02115, USA

1. *Nature* **405**, 608 (2000).

2. O'Reilly, M. S. et al. *Cell* **79**, 315–328 (1994).

3. Cao, Y. et al. *J. Biol. Chem.* **271**, 29461–29467 (1996).

4. Cao, Y. et al. *J. Biol. Chem.* **272**, 22924–22928 (1997).

When peer review fails

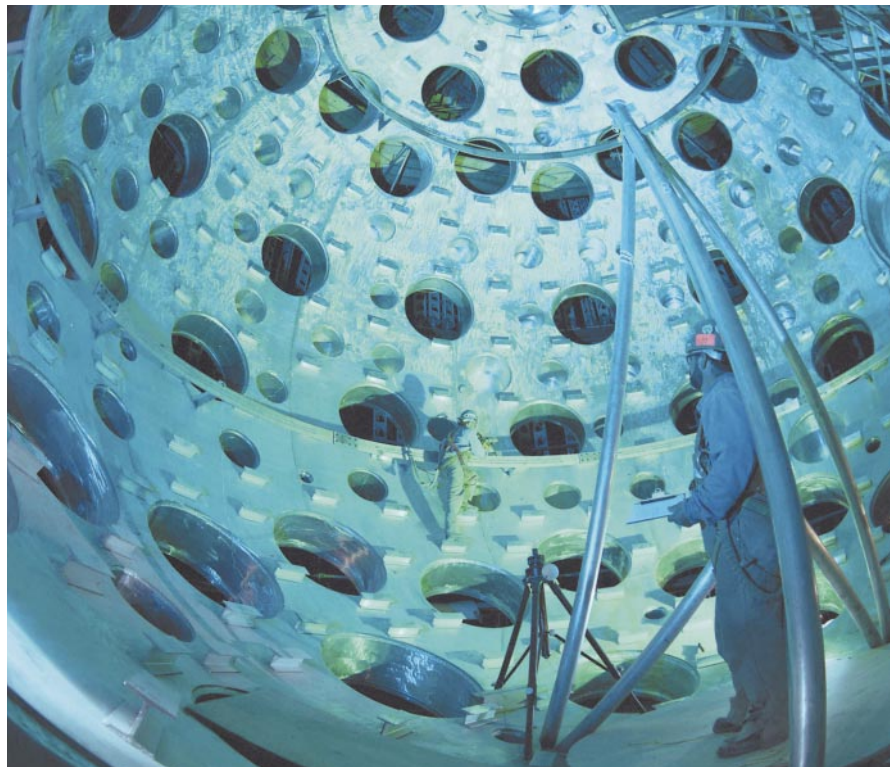
Problems with a giant laser project.

**Stephen Bodner and
Christopher Paine**

Given the numerous upbeat external reviews garnered over the years by the Lawrence Livermore National Laboratory's laser fusion project, the National Ignition Facility (NIF), its current problems (see box) may come as a surprise (see *Nature* 405, 110; 2000). But the reluctance of the US Department of Energy (DoE) and the Livermore laboratory to acknowledge and systematically explore underlying science and technology issues through independent peer review is an obvious contributor to the project's difficulties.

The US\$4 billion NIF project began life in 1990 as a proposed \$400-million glass laser. It was endorsed by the National Academy of Sciences (NAS) as "the only candidate laser driver that could be used for a [fusion] ignition demonstration in the next decade ... for a reasonable cost". When the United States halted underground nuclear tests in 1992, Livermore reworked its proposal for igniting tiny tritium-deuterium targets into a stadium-sized, 192-beam, \$1.1 billion project. The laboratory was aiming for a central role in the DoE's new "stockpile stewardship" strategy, in which data from above-ground experiments would be integrated with three-dimensional computer simulations to compensate for the loss of nuclear tests.

There were many outside reviews before construction of NIF began, but most were compromised by their lack of independence from Livermore and the DoE; a lack of balance among the panel members; and, on occasion, blatant conflicts of interest and "stacking" of committees. Last month, the General Accounting Office of the US Congress found that the reviews "did not discover and report on NIF's fundamental project and engineering problems, bringing into question their comprehensiveness and independence", and that even after senior project officials identified problems in late 1998,



Off target? The National Ignition Facility has massively overrun its original budget.

"there have been no effective independent reviews of NIF that have related technical challenges to its cost and schedule risks".

The first main review of the DoE's Inertial Confinement Fusion programme before NIF construction started was by the 1990 NAS committee. Chaired by Steven Koonin of the California Institute of Technology (Caltech), this established technical milestones in laser-target physics and laser development required before construction could begin. Los Alamos and Livermore scientists then discovered that the proposed target design was not viable. Livermore created a new design, but the technical milestones were not updated. Similarly, "Beamlet", a single-beam proof-of-principle

laser for NIF, was still being developed a year after ground was broken on NIF construction. When Beamlet ran into technical difficulties, Livermore simply announced success, dismantled it and continued building NIF.

Why was Livermore allowed to proceed on such a fast track, without first demonstrating all of the performance parameters required for ignition, using a full-scale, fully engineered prototype of a NIF beamline? In this instance the DoE was not an independent provider of funds in search of probing advice, as it had already decided that NIF would be the centrepiece of its new strategy. Unfortunately, it is a rare scientific review panel that tells the sponsor a programme is not ready to proceed when the sponsor strongly desires that programme, and when there is a hard sell by the lab conducting the project. The prevailing view became that NIF was critical to the "science-based" stockpile stewardship strategy, and to the future of Livermore laboratory.

The advocates became captives to their own rhetoric, and dissenting voices were ignored. For example, Seymour Sack, who designed and managed the development of nuclear weapons at Livermore for many years, was reported in *Science* (277, 304; 1997) as saying that the stockpile would not be any worse off without NIF. A former direc-

A project in crisis?

- The estimated cost to develop and complete construction of an "ignition demonstration" laser facility has risen from \$400 million in 1990, to \$1.2 billion in 1997, to \$1.6 billion in 1999, to \$2.2 billion in April 2000, and to \$3.9 billion (including ignition R&D costs) in August 2000 according to a General Accounting Office report (RCED-141), which warns that further R&D could drive the cost even higher (see <http://www.gao.gov>).
- The date for completing the full laser has slipped from 2004 to 2008.
- The routine operating energy for avoiding optical damage has been reduced to between a third and a half of the project's promised 1.8 megajoules, considerably less than what even the sponsoring laboratory believes is minimally required to achieve thermonuclear ignition.
- The project's primary justification is now shifting from fuel ignition to its residual capability for "high-energy-density physics" experiments, but this has never been seriously reviewed, and cheaper approaches are available.

tor of Los Alamos National Laboratory, Harold Agnew, is recently cited as saying that most specialists do not see the relevance of NIF to maintaining the active bomb stockpile (see http://www.abqtrib.com/archives/sports00/033000_nif.shtml).

Endorsing NIF

The lack of consensus resulted in intense manoeuvring to promote NIF. The next main committee to review the project was the DoE's Inertial Confinement Fusion Advisory Committee. In June 1995, one of us (S.B.) informed this committee of potentially crippling problems in NIF's target physics, and urged it "to alert the DoE to the current problems in the NIF target".

The upshot was a proposal by the chairman and several members to investigate these issues further. But before the next full committee meeting, Livermore NIF programme leaders wrote to the DoE criticizing the committee for its "very intense focus on the likelihood of ignition", and urging the replacement of six members. In September 1995, Victor Reis of the DoE discontinued the committee and replaced it with a new NAS panel. According to an internal DoE memo at the time, "a major review of the Inertial Confinement Fusion programme is needed in this fiscal year to reaffirm mission need and give further credence to arguments for success of the National Ignition Facility".

Koonin, by this time vice-president and provost of Caltech, was appointed chairman of this new review. This committee was seriously unbalanced: 11 of its 16 members had either previously stated positions supporting NIF and/or were consultants or advisers to Livermore or even the NIF programme itself. While serving on the committee, three members were directly involved in (successful) bids for closely related DoE contracts. Overall, 14 members had a personal or institutional connection with the very agency whose programme was supposedly undergoing independent review (see <http://www.nrdc.org/nuclear/rftr/rftrinx.asp>).

After the DoE and NAS rebuffed all informal attempts to foster a full and fair review, the Natural Resources Defense Council filed a suit in March 1997, obtaining a federal court injunction barring any DoE reliance on or dissemination of the committee's report, which had pronounced the NIF project ready for construction. Deprived of any official reliance on what had been billed as essential, the DoE forged ahead, breaking ground on the project in May 1997.

In summer 1999, Livermore managers surprised senior DoE officials with the news that NIF would have a cost overrun of between \$300 million and \$400 million, and a delay of a few years because of unexpected difficulty in assembling the laser in a clean environment, and too small a budget contingency. There was no mention of technical problems with either the laser or the target.

Two new review panels were established, one by the University of California (which manages Livermore) and the other by the Secretary of Energy Advisory Board (SEAB).

The chairman of the university panel, once again, was Koonin. This panel included scientists who had previously stated their support for NIF, and who had financial connections to the NIF programme. Its report was cursory, expressing optimism that any remaining technical problems could be solved.

Because the DoE was told that the difficulties were a management problem associated with the clean assembly of the laser, the membership of the SEAB's task force emphasized people with expertise in project management and clean rooms. The panel had only two laser experts, one who had served on a previous review supporting NIF, and the other a lower-ranking colleague of Koonin.

Even so, the SEAB task force discovered from an internal Livermore report that optical damage would limit the laser energy output initially to about 0.6 megajoules (see <http://www.nrdc.org/nuclear/nif/nif1104.pdf>). It concluded that aggressive research and development in materials science would be needed to raise the laser energy to one megajoule. Achieving the 1.8 megajoules needed for ignition experiments, while preventing the intense ultraviolet beam from damaging the final focusing optics, would now be a "challenging goal". The DoE panel supported Livermore's optimism that these goals could be achieved, but it was finally understood that the NIF laser remains, "at its core, a research and development project", not merely a construction project with management and clean-assembly problems.

Fundamental questions

These recent NIF reviews still suffered from basic failures in the peer-review process. Fundamental questions remain to be addressed by either an independent panel with the necessary expertise, or one that includes a fully representative cross-section of the critics.

With more funding and time, can the NIF project meet all the laser specifications required for its "baseline" ignition target design? If the laser could meet specifications, can the target design achieve ignition? Do the contributions of NIF to maintaining the nuclear weapons stockpile justify its cost, whether or not it reaches ignition? Is Livermore's entrepreneurial, risk-taking behaviour appropriate for an organization now responsible for maintaining rather than designing nuclear weapons? These questions are complex, and we have space here only to outline why we think the answer to all four is 'no'.

The optical damage problem is likely to limit NIF's usable energy for the foreseeable future to a megajoule or less, short of what is needed to evaluate ignition targets. But the performance of large glass lasers is often limited by more than optical damage: for example,

deterioration in other components, and non-linear optical effects that can destroy the quality of the laser beam, thereby affecting the final spot size. There were several indications of these additional problems during the Beamlet programme, but no laser shot ever simultaneously tested the full set of performance goals: full energy, full pulse-length, frequency converted and focused onto a target. Without such tests, one can only speculate on the beam energy that might be attainable on the NIF.

Recommendations

We recommend that Congress cap the current procurement and assembly of components, and make further NIF funding contingent on rigorous fulfilment of technical milestones using one complete 'bundle' of eight beams. Congress should not press ahead before Livermore has shown that the NIF laser system can meet all the project's technical requirements at tolerable cost.

There are sharp disagreements about whether any NIF-scale target can reach ignition, even if the laser operates perfectly. The physics of the laser-target interaction is complicated and controversial (and does not mimic the behaviour of nuclear weapons). There are also problems in the fabrication of the target coating surrounding the fusion fuel. A new panel should evaluate whether there is any real target design with the potential to achieve ignition. We do not believe there is, and the NIF project is far too expensive to gamble that viable targets will simply materialize when needed.

Irrespective of NIF's eventual capacity to achieve ignition, this facility cannot produce the same extreme physical conditions generated in a nuclear weapon explosion. Some scaling is required. It would be risky and unwise to rely on such extrapolations to evaluate future modifications to, or ageing of, nuclear weapons. A better strategy would be to ensure future capabilities for refurbishing and remanufacturing nuclear explosive components with minimal changes.

We are also concerned about Livermore's entrepreneurial, risk-taking ethos, its aggressively optimistic and even deceptive advocacy of NIF, and its refusal to admit the potential for laser problems. Scientists involved in the stewardship programme should be cautious, and willing to discuss problems with their government sponsors even when solutions are not available. Cautious scientists are not inferior to aggressively optimistic ones, and hasty construction of a flawed multi-billion dollar facility is not the way to attract the best scientists to the government's laboratories. ■
Stephen Bodner retired in 1999 as head of the laser fusion programme at the US Naval Research Laboratory (e-mail: a-s.bodner@mindspring.com). Christopher Paine is a senior researcher in the nuclear programme at the Natural Resources Defense Council, Washington, DC 20005, USA (e-mail: chrispaine@earthlink.net).

Sporting colours

Does skin colour have any true bearing on an athlete's ability?

Taboo: Why Black Athletes Dominate Sports and Why We're Afraid to Talk About It

by Jon Entine

PublicAffairs: 2000. 400 pp. \$25, £17.50

Kenan Malik

Saturday week, the Olympic men's 100-metre champion will cross the finishing line less than 10 seconds after leaving the starting-blocks. The winner may be the United States' Maurice Greene. He may be Trinidad's Ato Boldon. He may even be Britain's Dwain Chambers. But whoever takes the gold medal, one thing is certain: he will be black. Indeed, it's a racing certainty that all eight finalists will be black, just as they have been at the past four Olympics. Almost as certain is that every race from the 100 metres to the 5,000 metres will be won by an athlete of African descent — just as they were in Atlanta four years ago.

What lies behind such black domination of the track? And not just of the track. Sixty per cent of American footballers, for instance, are African American. Among basketball players, the figure rises to 80%. Of basketball stars, almost 95% are black. The traditional liberal answer points the finger at social factors. Blacks, so the argument runs, have been driven into sport because racism has excluded them from most other areas of employment. Racism also makes blacks hungrier than whites for success, and so they more often end up on the winners' rostrum.

American journalist Jon Entine dismisses this as "political correctness". The liberal consensus, he argues in *Taboo*, has served only to disguise the truth about the black domination of sport — which is that blacks are naturally built to run and to jump. There are two parts to the genetic story that Entine wants to tell about black domination. The vast majority of top-class long-distance runners are East Africans, and in particular Kenyans, mostly from the tiny Kalenjin tribe. Entine pulls together a number of lines of research which suggest that Kalenjin athletes are successful in the endurance events because they enjoy a slighter build than whites, have relatively longer legs and larger lung capacities, and because their muscles have a higher proportion of slow-twitch fibres and so are better able to utilize oxygen.

By contrast, Entine argues, athletes of West African descent dominate in "such anaerobic activities as football, basketball and sprinting" because they have the most mesomorphic physiques, with "bigger, more visible muscles including a larger chest". They possess less body fat, a higher centre



In a winning frame? Maurice Greene's victory may owe much to his mesomorphic physique.

of gravity, narrower hips, higher levels of plasma testosterone, and a higher proportion of fast-twitch muscle fibres.

For Entine, such physiological and biomechanical differences demonstrate the natural superiority of black athletes. For Entine's critics, on the other hand, the very search for such differences demonstrates a racist outlook. "Some things are better left unsaid," claimed *The New York Times*. Entine acknowledges that this kind of research could be used for racist ends. Some controversial scientists, such as J. Phillippe Rushton, claim that there is a trade-off between brain and brawn, and that black athletic superiority has been purchased at the price of lower intelligence. But, as Entine rightly argues, however offensive such claims may be, the cause of anti-racism is not strengthened by ignoring science or censoring data. The real trouble with *Taboo* is not that it treads on dangerous ground, but that its thesis that white men can't jump (or run) does not convince.

The most basic problem is Entine's confusion of racial and population differences. It is true that gene frequencies vary across populations, so that some populations may be more athletically talented than others. But this is not the same thing as saying that blacks are "born to run". As population geneticists never tire of telling us, 'race' has little biological validity. It is certainly possible to divide humanity into a number of races, as we conventionally do, according to skin colour and body form. But it is also possible to do it in many other ways — using, for instance, blood group, lactose tolerance, sickle cell, or

any other genetic factor, as the basis for our new 'races'. Genetically, each would be as valid a criterion as skin colour. The current division of the world into black, white and Asian races is, in other words, as rooted in convention as in genetics.

Entine rejects such criticisms as mere 'semantics'. But his own argument shows why it is not so. According to Entine, East Africans are naturally superior at endurance sports, West Africans at sprinting and jumping, and "whites fall somewhere in the middle". But if East and West Africans are at either end of a genetic spectrum of athletic abilities, why consider them to be part of a single race and one that is distinct from whites? Only because conventionally we use skin colour as the main criterion of racial difference.

Entine also ignores evidence that does not fit his thesis. It is true that athletes of West African descent living in North America, Western Europe and the Caribbean dominate many sports. But contemporary West Africans don't. This is the opposite of what one should expect if athletic ability was predominantly genetic. In the United States, considerable intermixing between black and white populations has meant that the African-American population embodies, on average, some 30% of 'white' genes. Hence African Americans should be poorer athletes than West Africans. The reverse is true.

Entine suggests that West Africans are built to jump better than whites. But in the three jump events — high jump, long jump and triple jump — blacks do not dominate whites in the way they do on the track. The

book reviews

pole-vault — an event that requires both sprinting and jumping ability — is almost entirely white, as are field events such as discus, javelin and shot-put.

Entine points out that a cluster of South Pacific islands, such as Samoa and Fiji, have in recent years produced an extraordinary number of American football and Australian rugby players. Such islanders, Entine argues, “tend to be large and explosively fast”, implying that their success somehow gives credence to his ‘blacks are bred to run’ thesis. In fact it does the opposite: genetically, there are few populations in the world more different from sub-Saharan Africans than Pacific islanders. For whatever reason Fijians and Samoans excel at sport, it is not because they genetically resemble Nigerians or Kenyans.

The most irritating aspect of *Taboo* is Entine’s constant dismissal of critics — who include the eminent scientists Richard Lewontin and Luigi Luca Cavalli-Sforza — as “postmodernists”, or simply motivated by ideology. Such bluster cannot hide the gaping holes in his argument. Is athletic talent at least in part inherited? Undoubtedly. Are there genetic differences between populations? Clearly. Are West Africans and Kenyans genetically built for running? Possibly. Are blacks naturally better athletes than whites? Not necessarily. After all, how many African pygmies have you seen garlanded with Olympic medals? ■

Kenan Malik is a freelance writer living in London, UK (e-mail: kenanmalik@msn.com).

At the edges of life

Life at the Extremes: The Science of Survival

by Frances Ashcroft

HarperCollins: 2000. 326 pp. £17.99, \$24.95

Knut Schmidt-Nielsen

When we venture to the edge of livable conditions, or even beyond, the results are not always pretty: altitude sickness, diver’s disease, frostbite, heat stroke and so on. Frances Ashcroft, who describes such nasty events in this book, is a venturesome woman. Preceding each of her first four chapters is a tale of her own encounters with the effects of altitude, diving, heat and cold. Further chapters discuss the ultimate physical performance needed in competitive sports, and the physiological and technical problems of space travel. There is also a look at the extraordinary organisms that thrive in conditions beyond what any human would dream of tolerating.

Ashcroft gives much interesting historical information which helps us understand the dangers faced by mountain climbers, early balloonists and divers, many of whom paid for their experiences with their lives. She also gives fascinating accounts of how our knowl-

From one extreme to another

Two years after having brain surgery and chemotherapy for stage-four testicular cancer, US cyclist Lance Armstrong won the 1999 Tour de France. His autobiography, *It’s Not About the Bike: My Journey Back to Life*, written with Sally Jenkins (Yellow Jersey Press/Putnam; £17/\$24.95), races through his childhood in Texas, the chivalric code of the peloton, his existence during chemotherapy and his emergence as both a strategist and a sprinter fit to conquer time trials and mountain routes alike in the 1999 Tour. He was left for dead in a cancer ward by his then sponsors, Cofidis, only to be tipped by Miguel Indurain, winner of five consecutive Tours, to win the 1999 Tour. He met his wife while setting up the Lance Armstrong Foundation for cancer patients and survivors. The book is a raw and gripping read written in pure testosterone — quite an achievement for a woman ghost writer.



edge of human physiology has progressed and now forms the basis for our understanding of the limits to human accomplishment.

We can follow the superhuman efforts of the men who reached the 8,848-metre summit of Mount Everest. Many climbers died trying to conquer the world’s highest peak, suffering from the lack of oxygen and freezing temperatures. The top was finally reached in 1953 with the aid of extra oxygen; amazingly, it has since been scaled without this.

The increased pressure encountered in diving has serious dangers. Whether we use scuba equipment or a diver’s suit, the gases we breathe are at the same pressure as the surrounding water, and with increasing pressure as we dive deeper the gases we normally breathe, oxygen and nitrogen, become toxic and narcotic. Worse, they become dissolved in blood and body fluids, and when we return to the lower pressure at the surface, the dissolved nitrogen bubbles out of solution, causing the painful and often lethal condition known as diver’s disease, or the bends.

If humans suffer from the bends after diving beyond some ten metres, how can whales and seals avoid them when diving to more than 1,000 metres? Simple. When they begin a dive, they exhale, so there is no gas in their lungs to enter the blood and cause trouble when they surface.

Our tolerance of heat is limited, and without enough drinking water, a single day in a hot desert is likely to end in a miserable death. But in 1775 the Secretary of the Royal Society, Mr Blagden, had a room heated to well above boiling point, and, together with a dog and a piece of raw meat, he entered the room at 105 °C. When he left it 15 minutes later, he and the dog were fine, but the steak was well done.

Ashcroft describes well how humans cope with cold, and her accounts of polar expeditions give us an insight into the limits of human endurance. The conditions that some polar explorers have suffered are astounding. A few photographs show effects of frostbite so appalling that one almost wishes they were as poorly reproduced as some of the other photos, which are so dark as to be indecipherable.

In the chapter on space, I learned that a space suit weighs a formidable 113 kilograms, although in space, of course, it is weightless. We also learn that a sudden decompression in space would cause immediate death as all air rushes out of the lungs, gas in the intestine expands explosively, and blood and body fluids boil — a grim but swift end.

We learn in the last chapter about the peculiar organisms that live around the black smokers in the deep ocean, where superheated water laden with hydrogen sulphide emerges. Strange invertebrates, known nowhere else, depend on bacteria that gain energy from the oxidation of the otherwise toxic hydrogen sulphide. We also meet other creatures that endure, and even prefer, extreme conditions.

The book is generally well and clearly written at a level suitable for non-physiologists. Ashcroft is at her best when she explains what happens to humans under extreme conditions. These explanations are easy to grasp, and the author only occasionally falls into the trap of using scientific jargon. For example, in a book like this there is no need to show the complex chemical structure of adenosine triphosphate, or to explain that Heisenberg’s uncertainty principle means that all molecular motion does not cease at absolute zero. The book is much too readable

and interesting to load it with such minor excursions into scientific technicalities. ■

Knut Schmidt-Nielsen is in the Department of Biology, Duke University, Durham, North Carolina 27708-0338, USA.

Surfing the nebulae, churning the waters

From Galaxies to Turbines: Science, Technology and the Parsons Family

by W. Garrett Scaife

Institute of Physics: 1999. 400 pp. £35, \$45

J. L. Heilbron

In 1845 Dean Peacock of the Church of Ireland strutted in top hat, with umbrella unfurled, through a wooden barrel 46 feet long. His parade was part of the inauguration of the Leviathan of Parsonstown, the six-foot mirror that, together with the barrel, made up the world's biggest telescope. Two masonry walls, each 50 feet high and 70 feet long, supported the instrument on the grounds of Birr Castle, the seat of the Earl of Rosse in County Offaly, central Ireland. The designer and builder of the telescope was the Earl's son and heir, William Parsons.

In 1897 the first turbine-driven ship, the *Turbinia*, demonstrated its unique speed and efficiency at the jubilee regatta held in honour of England's Queen Victoria. The *Turbinia* had not been invited. She brazenly sailed in and out among the great ships, easily eluding the torpedo boats assigned to intercept interlopers. The bold sailor and self-promoter, the inventor of the marine turbine, was Charles A. Parsons, the tenth child and youngest son of the creator of the outsized telescope.

William built as much of his telescope as he could in workshops around Birr Castle staffed by local men he had trained as mechanics and opticians. He encouraged all his children, the girls as well as the boys, to experiment in the workshops. His wife Mary, whose fortune helped to defray the astronomical cost of the Leviathan, also spent time there and became an adept photographer. Charles's eldest brother Laurence, who inherited the estate, continued the work on the telescope; but whereas William had built the instrument to study the structure of the most distant objects then known, the nebulae, which he did to general approval, Laurence used it to take the temperature of the nearest, the Moon, which met with undeserved coldness in the academic establishment. Nonetheless Laurence followed his father into the Royal Society; Charles also became a fellow. One of their brothers had a successful career as an engineer. Various Parsons nephews and relatives of William's workmen also had talents for technology and found places in Charles's enterprises.

Although Charles attended Cambridge, from which he emerged eleventh wrangler in 1877, he followed his father's style of invention: by inspired cut-and-try. Just as William had experimented tirelessly with the make-up of speculum metals and the techniques of pouring, cooling, annealing, grinding and polishing them, Charles varied the number, orientation and shape of the turbine blades, the velocity of the steam through the turbine, the lubrication and placing of the bearings, and so on, until he had maximized performance. Modern modelling confirms that the propulsion package of the *Turbinia* was near optimal for the ungeared drive system Parsons invented.

Few nineteenth-century Oxbridge men and still fewer members of the landed aristocracy were willing to labour for days with workmen, to endure physical hardship or to neglect personal comfort in order to make mirrors smoother or rotors better. Garrett Scaife credits the Parsons' obsession, or anyway its indulgence, to their avoidance of English public schools. In Ireland one could be a lord and a gentleman while also devoting one's time to manual work.

The family obsession and creativity appeared most obviously in Charles, to whom Scaife devotes most of his book. As the youngest son, Charles felt he had to work for a living. A measure of his mettle is that, after losing his first patents on the turbine to partners from whom he had parted, he did not think to quit or to compromise. He invented another type of steam turbine, and no doubt would have persisted with it had his former partners not changed their minds and sold him the original patents at less than 2% of what they had originally demanded. With this sort of grit came irascibility and even irrationality. But Parsons could also be gentle and generous when alerted to the troubles of people around him. Although often in the company of distinguished men, he was not at ease in society. His wife Katherine, the tenth child and youngest daughter of landed gentry in Yorkshire, recalled that when they met "he had the character of being an extraordinary and weird young man socially but it was understood that he was a genius".

The "genius" invented or improved many things besides steam turbines. He began with an unsuccessful rotary engine, went on to torpedoes, designed electrical apparatus, found a way to mould parabolic mirrors, failed to make diamonds, and did much else. But propulsion was his compulsion. He devoted himself with such energy to perfecting the marine steam turbine that only a decade after *Turbinia* had demonstrated its feasibility he began to build a propulsive plant with 16 times its power. The result of this scale-up was designed to drive the *Mauretania*, which would be the largest ship in the world.

Parsons did not care to monopolize his businesses or drive hard bargains. He paid

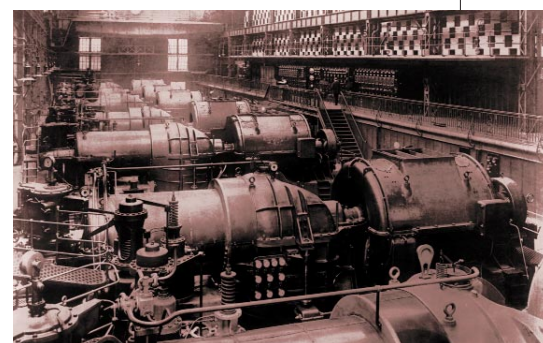
himself no salary, obtaining from the works enough to maintain his development programmes and satisfy his shareholders. His main backers, aristocrats enlisted by Laurence and wealthy men interested in the technical progress of the turbine, gave him a free hand in choosing his projects, staff and business.

He was not a good manager, but was astute enough to act on suggestions from people who were. He gave one of his nephews, who had learned industrial management at Westinghouse, responsibility for introducing US-style efficiency. The nephew's brief did not extend to machinery. In the 1920s engineers from a big shipbuilder in Hamburg came to see Parsons' shops. They were impressed: "It is wonderful how you achieve such accuracy with such old-fashioned machinery." A programme of retooling followed.

Constant work and an irascible temper did not make Charles a good family man. His wife spent her time improving the huge estate Charles bought to indulge his passion for hunting. The First World War made her a champion of women's rights in the workplace. Many women learned to be excellent mechanics when their men went off to fight, but at the war's end most were let go. Parsons' works participated in this general pattern, with the notable complication that one of the women engaged in a senior position was his daughter Rachel, who had studied engineering at Cambridge. She soon left for a government job and later had a position in a factory set up by her mother and run by women. The only son of the family was killed at the front.

Parsons' engineering works barely survived the war. But with the help of friends and the Second World War it came to prosper again. In 1953 it received a contract for the generators of the world's first nuclear power plant. For a few decades more it carried on as the only large manufacturer in its field not owned by a foreign firm. This happy state ended in 1997, when Siemens acquired "C. A. Parsons Works". They have retained the name.

Parsons' business brought him into frequent contact with officers of the Royal Navy and with politicians and other administrators. Scaife supplies anecdotes about their ignorance of technical matters. Parsons often



Energetic applications: Charles Parsons' steam turbines were used in the Carville power station.

spoke out for inventors and for technical education. Unfortunately, however, much of the Parsons tale is from a time now gone. The heroic inventor with a small staff, choosing his problems as he pleases, obsessed with perfection, creative, honest, indifferent to riches — such as Parsons, Thomas Edison and Polaroid inventor Edwin Land — may be no more. The quick fix, the fast buck, the multinationals and the prevailing greed make their birth difficult and their survival unlikely.

From Galaxies to Turbines is stuffed with facts, figures, diagrams, pictures and quota-

tions, many from archival sources. The pile of details helps to measure the breadth and height of the Parsons' achievements. These have an authoritative analyst in Scaife, formerly a power engineer, lecturer in mechanical engineering and fellow of Trinity College, Dublin. He could have let himself go even further than he has, for some diagrams need more explanation than they are given. But it is not necessary to follow all the details. You do not need them to freeze vicariously with William and his friends on clear, cold, winter nights looking for nebulae, or to broil near

the furnaces where he melted his mountains of glass. With a little attention you can shiver with Charles in the spray and sicken with the swells as *Turbinia* suffers through her many trial runs, or feel the fatigue and frustration of his incessant tinkering. Scaife's unpretentious text makes plain the courage, stubbornness and stamina that created the Parsons family Leviathans and offers welcome refreshment among the gimcracks and glitter of more fashionable histories of science. ■

J. L. Heilbron is at Worcester College, Oxford OX1 2HB, UK.

Science in culture

Art advances science

Today's scientists stand on the shoulders of pioneering artists

Robert S. Root-Bernstein

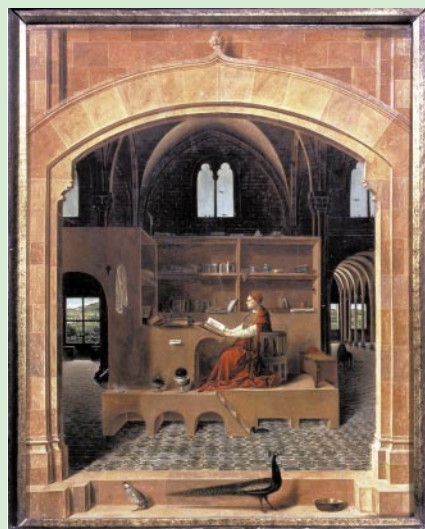
At the Nobel conference in 1980, director Thomas Gover asked his colleagues: "Can you think of an example where an artist has supplied the missing piece in some understanding of the physical world?" William Lipscomb, a laureate in chemistry, replied: "Not an original one." Escher, he suggested, may have added something to the scientific exploration of colour symmetry. The physicist Freeman Dyson offered Goethe's colour theory, cautioning that it "turned out to be a rather dismal failure". Art, the participants concluded, has provided nothing to science.

I hesitate to disagree with men as accomplished as Dyson and Lipscomb, but they are wrong. The arts often contribute to modern science. While space permits only a few exemplars, these contributions can be found in every science and include the invention of new structures, techniques and aesthetic sensibilities.

Artists often invent new structures that scientists then discover in nature. Virologists attempting to understand the structure of the protein shells that surround spherical viruses such as polio during the 1950s were directed by knowledge of Richard Buckminster Fuller's geodesic structures. These also became the models for numerous carbon molecules aptly named fullerenes, including the perfect geodesic dome C_{60} — buckminsterfullerene.

Sculptor Kenneth Snelson has had a similar impact on scientific thinking through his role in the invention of tensegrity, a principle by which a stable structure is created by linking stiff, non-compressible units (such as rods) with highly flexible material (such as a rope) under tension. Donald Ingber and Steven Heidemann recognized that tensegrity sculptures have many similarities with protein structures and have been modelling them using this art.

Artist Wallace Walker, while studying in the 1960s, was asked to make a three-dimensional object out of a sheet of paper only by folding and gluing it. The result was a complex doughnut that could be folded through its centre hole into a kaleidoscopic variety of shapes. Doris



Renaissance perspective painting, as seen in *St Jerome in His Study* by Antonello da Messina.

Schattschneider, a mathematician specializing in geometric objects, determined that Walker's paper sculpture was the first of a novel class of geometric objects, now called kaleidocycles.

Many scientific techniques also originate in art. Anamorphosis — 'shape change' — derived from the Renaissance discovery of perspective drawing: mapping a three-dimensional object onto a flat surface. Artists then realized that two-dimensional objects could be mapped onto three-dimensional surfaces, including spheres, cones and rods. Such transformations became central to D'Arcy Thompson's *On Growth and Form* and Julian Huxley's *Problems of Relative Growth*, both of which describe evolutionary and embryological processes as anamorphic distortions. Anamorphosis also underlies Wilder Penfield's and Clinton Woolsey's studies of the motor and sensory mappings of primates onto the cortex of their brains, which yield homunculi with huge lips, hands and feet, and tiny bodies.

Another striking example is the reification of logic in modern computer chips. Chips are manufactured using methods adapted directly from silk screening and etching. Logical operations are carried out in electronic gadgets only because the art exists to transform them into

physical patterns, and these patterns exist only because their designers understand how to transform logical operations into images.

Finally, the arts can foster scientific advances through the development of new aesthetics. The process of breaking a picture into discrete areas of colour (pixels) was invented by pointillist painters such as Seurat. The technique of false-colouring objects, which scientists use to emphasize inobvious elements of data, was invented by Fauvist painters. Abstract art, in which a single element of a complex phenomenon (such as its pattern, structure or colour) is chosen for selective description, was pioneered by Picasso and Kandinsky in the 1920s.

Indeed, there is growing understanding that art fosters science. Mitchell Feigenbaum, one of the pioneers of chaos theory, believes that understanding how artists paint will provide the cognitive insights necessary to do better science. "It's abundantly obvious that one doesn't know the world about us in detail," he said. "What artists have accomplished is realizing there's only a small amount of this stuff that's important, and then seeing what it was. So they can do some of my research for me." C. S. Smith of MIT spent a lifetime studying oriental arts and crafts for the insight they gave him into metallurgy. "I have slowly come to realize that the analytic, quantitative approach I had been taught to regard as the only respectable one for a scientist is insufficient," he said. "The richest aspects of any large and complicated system arise from factors that cannot be measured easily, if at all. For these, the artist's approach, uncertain though it inevitably is, seems to find and convey more meaning."

In short, the arts and sciences are as integral today as they were in the Renaissance. We must foster the connections and the people who can make them. For as Robert Mueller, an MIT-trained engineer and artist, wrote in his stimulating book, *The Science of Art*, "Art may be a necessary condition for constructing the new consciousness from which future science gets its structural realities to match nature, in which case it is more important than we generally admit." ■

Robert S. Root-Bernstein is in the Department of Medical Humanities, Michigan State University, East Lansing, Michigan 48824, USA.

Creative tension

What links Aristotle, William Blake, Darwin and GM crops?

Keith G. Davies

In 1809, at 28 Broad Street, Soho, in London, there was an exhibition of paintings by the then-neglected artist William Blake, including "A Subject from Shakspear ... The Horse of Intellect is leaping from the cliffs of Memory and Reasoning; it is a barren Rock: it is also called the Barren Waste of Locke and Newton". Blake loathed reductive science and the mechanized materialism of the Industrial Revolution. His Aristotelian way of thinking has a history dating back to Plato.

Aristotle believed that although each individual died and perished, its ideal 'Form' was eternal and fixed as a species. Or, as Blake put it: "Whatever can be Created can be Annihilated: Forms cannot: The Oak is cut down by the Ax, the Lamb falls by the Knife, But their Forms Eternal Exist For-ever." Charles Darwin was less than a year old at the time of Blake's exhibition, and it would be 50 years before his reductions would take on Aristotle's holistic thinking. Yet the current furore over genetically modified organisms

is one modern consequence of the ancient conflict between these two ways of thinking.

In the *Origin of Species*, Darwin presented his idea of natural selection in terms of analogies to the tradition of artificial selection on plants and animals. Darwin replaced Aristotle's concept of species as eternal, ideal Forms with one based on groups of individuals in a population. Darwin's population thinking stresses the uniqueness of everything in a related living world, where the species is a statistical abstraction and only unique individuals have a reality. Aristotle thought the reverse: the ideal Form is real and individual variation is an illusion. Darwin's view is reductionist and individualistic, Aristotle's holistic. As Ernst Mayr has said, "No two ways of looking at nature could be more different."

After Darwin, holistic thinking had to change its form to survive: although Darwin had shown that species were mutable, a debate raged into the next century about whether life could be reduced to physics and chemistry. As Niels Bohr argued, "Vitalism scarcely finds its proper expression in the old supposition that a peculiar vital force, quite

unknown to physics, governs all organic life ... if we were able to push the analysis of the mechanism of living organisms as far as that of atomic phenomena, we should scarcely expect to find any features differing from those of inorganic matter."

During the Second World War many physicists who were driven to address vitalism entered Britain and the United States from mainland Europe. A research effort began to try to understand how life begets life, but not quite. We call the result molecular biology. But what about vitalism and holistic thinking? Can biology be reduced to just physics and chemistry? The answer is subtle.

In physics and chemistry, things tend to equilibrate: but for living cells equilibrium means death. Molecular biology was born with the recognition that the large molecular structures essential to living cells gave rise to autonomous feedback systems. They are based on chemical circuitry and yet, as François Jacob put it, they "transcend chemistry". The living and the non-living do not differ in composition, but rather in their level of organization. All life can be seen as an interacting hierarchy of feedback systems, from cells to individuals, and from populations to communities and ecosystems.

Aristotle and Plato believed that humans could use "intellectual intuition" to visualize eternal species and discriminate between them. Given that this view of immutable species lasted for more than 2,000 years, it is not surprising that many people today find the mere thought of taking a gene from one species and placing it in another abhorrent.

Aristotle faced strong opposition even in his time: a contemporary, Antisthenes, is reported to have said: "I can see a horse, but I cannot see horseness." But the Aristotelian/Platonic view was doctrinal in the West until Darwin.

There is still tension between the holistic ideal and the material reductionist way of thinking; this tension maintains openness and is progressive. For science to have a healthy future, the balance between these approaches must never become dogmatic. Our imagination gives our guesses a holistic basis, our reductive experiments a way to falsify them: the confrontation is essential. Or as Blake put it, "The true nature of knowledge is experiment," but "what is now proved was once only imagin'd."

Keith G. Davies is in the Entomology and Nematology Department, IACR-Rothamsted, Harpenden, Hertfordshire AL5 2JQ, UK.



The Genius of Shakespeare by William Blake: Blake's holistic thinking, which originated in Aristotle's time, would be challenged from Darwin onwards.

Madame Bovary, c'est moi

Once upon a time, fiction required the suspension of disbelief.

Dan Simmons

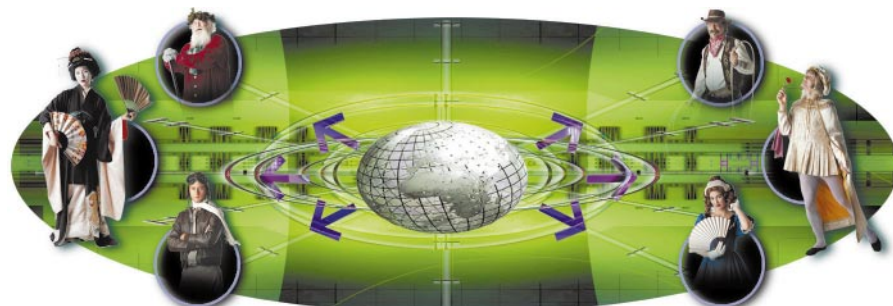
2052 AD: the current migration of millions of people to novel universes by quantum teleportation (QT) was a shock even to the system's inventor, Jian-Wei Martini. "I don't know why it took me by surprise," said Dr Martini in a 2043 interview. "I had the basic idea for QT when I read an ancient *New Yorker* story by Woody Allen, but the potential was all there in Schrödinger's classic wave equation." Dr Martini has since QT'd to Madame Bovary's universe and is currently living in Flaubert's Paris as Monsieur Leon.

The quantum-mechanical entanglement effect had been analysed (in Einstein's sceptical term) as "spooky action at a distance" since 1935, but it was not until 1998 that a research group at the University of Innsbruck demonstrated actual quantum teleportation of a photon — or more precisely, the complete quantum state of that photon.

These initial quantum-state teleportations avoided violating Heisenberg's principle and Einstein's speed-of-light restrictions because teleported photons carried no information, even about their own quantum state. However, by producing entangled pairs of photons and teleporting one of the pair while transmitting the Bell-state analysis of the second photon through subluminal channels, the recipient of the teleported photon-data had a one-in-four chance of guessing its quantum state and then utilizing the quantum bits of teleported data.

All of this would have amounted to very little except for remarkable advances in human-consciousness research. Researchers at the University of Kiev interested in improving memory function used quantum computers to analyse biochemical cascades in human synapses. In 2025, they discovered that the human mind — as opposed to the brain — was neither like a computer nor a chemical memory machine, but exactly like a quantum-state holistic standing wavefront.

The human brain, it turned out, collapsed probability functions of this standing wavefront of consciousness in the same way that an interferometer determined the quantum state of a photon or any other wavefront phenomenon. Using terabytes of qubit quantum data and applying relativistic Coulomb field transforms to these mind-consciousness holographic wavefunctions, it was quickly discovered that human consciousness could be quantum-teleported to points in space-time where entangled-pair



wavefronts already existed.

And where were these places? Nothing as complex as an entangled consciousness-wavefront existed elsewhere in our continuum. QT researchers soon realized that they were teleporting human consciousness — or, rather, the complete quantum state of these consciousnesses — to alternate universes that had, in turn, come into being through the focus of pre-existing holographic wavefronts: in other words, complete alternate universes created by the sheer force of human imagination. These singularities of genius act as Bell-state analyser/editors on the quantum-foam of reality, simultaneously interpreting one universe while creating a new one.

It seems that poets, playwrights and novelists already understood this. "The imagination may be compared to Adam's dream," wrote John Keats: "he awoke and found it truth."

The QT charting of 'fiction universes' began immediately, but even before a hundred alternate universes were confirmed, the QT migration from our Earth began.

Meanwhile, the 'agon' — the imperative to rank the relative importance of creative works — now had a scientific tool at its disposal. The long literary debate as to which works belonged in the so-called 'Western Canon' was settled by QT exploration. For example, 21 of Shakespeare's 38 plays generated complete alternate universes, as complex and expansive as our own, each capable of supporting a human population from a few thousand (*Measure for Measure*) to many billions (*King Lear*), despite the fact that each play may have had a cast of a score or fewer players when it was performed. More than a million people have migrated to Elsinore, whereas fewer than 5,000 — mostly clinically depressed Scandinavians — have seen fit to re-establish themselves in Lear's universe.

Flaubert, it turned out, generated two complete universes — the so-called "Madame Bovary's World" and that of *Senti-*

mental Education — whereas Alice Walker, it seems, to the frustration of American academics, had created none.

The alternate universe of Dante's *Inferno* has received more than 385,000 emigrants (mostly from Southern California), but his Paradiso Planet shows only 649 transplants. The current count of those who have QT'd to Huckleberry Finn's "River World" is 3,622,406, and more than a million have been transported to each of Charles Dickens's five extant universes.

It is true that more than 60 million people volunteered to QT to D. H. Lawrence's "Lady Chatterley's World", but — sadly — no such discrete universe has been found. Some have made do with the universe of *Sons and Lovers*. Recent QT stampedes to the worlds of Jane Austen, Robert Louis Stevenson and the effectively infinite number of universes created by Jorge Luis Borges may reflect current social trends.

It remains to be seen whether QT technology will solve the current global population crisis or if it will remain an option of the rich, bored and educated. It also remains to be seen whether any universes will be found — beyond the paltry handful already discovered — that owe their origins to 21st-century imaginations.

"That which is creative must create itself," said John Keats. And perhaps more pointedly, from William Blake — "I must create a system or be enslaved by another man's."

Which brings us to the central question of this new quantum reality. None of the millions of ancillary characters living in Madame Bovary's universe (except for the QT emigrés) know that they are characters in a work of fiction. Nor do they know that Madame Bovary is the main character.

So who wrote our universe? And who are the central characters?

Dan Simmons is the author of 16 books, four or five of which have been science fiction. His most recent novel is *Darwin's Blade* (published in October 2000 by William Morrow/HarperCollins).

Currents without borders

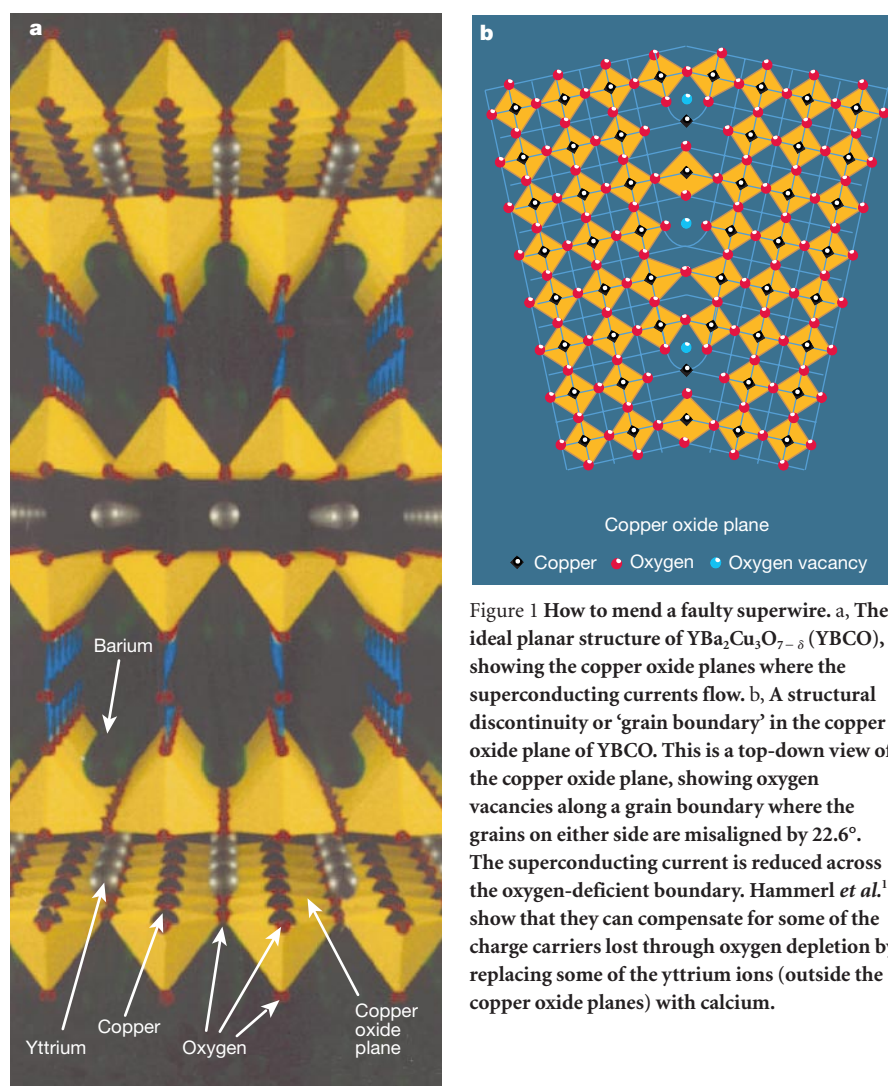
Paul M. Grant

Following the discovery of superconductivity at liquid-nitrogen temperatures, the idea of making 'superwires' soon ran into problems. Structural impurities remain the main obstacle, but a high dose of calcium may be the answer.

In the 1960s, a solid-state physics researcher who studied a non-cubic compound that was made up of more than two chemical elements would be committing professional hara-kiri. Not even oxides were considered fashionable. Such materials were the province of the 'dirtier' derivatives of the field — metallurgy, geology, polymer chemistry and, at best, crystallography. But by the 1970s we purists had run out of simple structures. Much to our surprise, we found an even greater richness of physics in conducting organic materials, amorphous metals and semiconductors, and multi-element transition-metal oxides.

No event better illustrated this epiphany than the 1986 discovery of high-temperature superconductivity in barium-doped lanthanum copper oxide, $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$. The process of doping (replacing some La^{3+} by Ba^{2+}) introduces positive charge carriers (holes) into the material that can flow without resistance under certain conditions. The transition temperature at which this 'superconductivity' happens, T_c , is much higher in the copper oxides than in other compounds (as high as 135 K) — bringing closer to reality the dream of superconducting wires that operate at liquid-nitrogen temperature (77 K), rather than liquid-helium temperature (4 K). Nowadays, not just high- T_c superconductors, but also their offspring — magnetoresistive compounds used in computer hard drives — have taken centre stage.

The material described by Hammerl *et al.*¹ on page 162 of this issue continues this love affair with more complex compounds. The authors' samples contain five elements in unequal amounts, and, horror of horrors, are built from messy multilayer thin films. Such polycrystalline compounds usually consist of multiple grains whose boundaries reduce the supercurrent flow. Could anything be less 'fundamental'? Yet it was in a similar structure that Kirtley and Tsuei² discovered the underlying symmetry of the superconducting ground state, a critical clue to the mechanism behind high- T_c superconductivity. Could anything be more 'fundamental'? Nonetheless, insulating grain boundaries in superconducting films severely hinder the flow of supercurrents over any appreciable distance involving multiple grains. This is a serious obstacle to the production of longer wires and tapes for electric power devices. Hammerl *et al.* think they



may have a clever structural solution.

All superconducting copper oxides have, without exception, highly anisotropic crystal structures (Fig. 1a). The superconducting state is most robust along the CuO_2 layers (the copper oxide planes), and it is relatively easy to grow polycrystalline thin films of the high- T_c compounds with these layers in the same plane as the film. Unfortunately, such films also contain a large number of grain boundaries where the grains on either side are misaligned by a large angle (more than 10°) (Fig. 1b). It has been shown that these large-angle grain boundaries form insulating 'Josephson junctions' — weak links that drastically reduce supercurrent

Figure 1 How to mend a faulty superwire. a, The ideal planar structure of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (YBCO), showing the copper oxide planes where the superconducting currents flow. b, A structural discontinuity or 'grain boundary' in the copper oxide plane of YBCO. This is a top-down view of the copper oxide plane, showing oxygen vacancies along a grain boundary where the grains on either side are misaligned by 22.6° . The superconducting current is reduced across the oxygen-deficient boundary. Hammerl *et al.*¹ show that they can compensate for some of the charge carriers lost through oxygen depletion by replacing some of the yttrium ions (outside the copper oxide planes) with calcium.

flow between grains, especially in an applied magnetic field³. All high- T_c compounds suffer from this problem to some extent, but it is especially pernicious in one of the compounds best suited for making superconducting wire, $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (YBCO), which has a T_c of 91 K. Calculations indicate^{4,5} that these grain boundaries have fewer charge-carrying holes, thereby reducing the ability of the boundary to conduct currents. Experimental data suggest^{6,7} that this effect might result from oxygen ions being squeezed away from the boundary by compressive strain.

What can be done? One obvious way to 'heal' oxygen depletion at the boundary, by heating the films in an atmosphere of oxygen



100 YEARS AGO

Another of those disastrous hurricanes which occasionally visit the West Indies and United States at this season of the year has to be recorded. On the 8th Inst. a storm of great violence struck the coasts of Louisiana and Texas, and, owing to the thickly populated districts over which it swept and to the high water wave which accompanied it, immense destruction to property and lamentable loss of life ensued. The fury of the storm is said to have been felt for at least a hundred miles inland, but up to the present time scarcely any details have arrived as to its character and the exact path that it followed. This part of America is one of the three regions referred to in the works of Prof. W. M. Davis from which tropical storms move into temperate latitudes in the northern hemisphere; but we must wait for further details before it can be stated whether the one in question was of the nature of a tornado, which differs from an ordinary hurricane chiefly in its excessive violence over a small, instead of a large, area. From the description so far given, and from its duration, the storm would appear to have been of the nature of the worst West India hurricanes.

From *Nature* 13 September 1900.

50 YEARS AGO

Publication of the Report of the Royal Commission on Population has been followed by a series of papers, and the fifth volume... contains an important contribution on "The Economic Position of the Family"... The authors of this paper, after discussing the various ways in which parents meet the extra cost of bringing up children, by comparing their expenditure with that of childless couples having the same income, conclude, first, that at all income-levels parents have to make considerable economic sacrifices to maintain their children, and, secondly, that children in large families have a lower standard of living than children in smaller families... It would seem that, despite the large increase of prices and incomes, the actual money cost of a child to its parents, at a low working-class level of income, is substantially unchanged as compared with the pre-war figure. The burden of two children, which at this income-level was about a third of a childless couple's income, has now fallen to one-sixth. The raising of the school-leaving age, however, has meant that the burden lasts a year longer.

From *Nature* 16 September 1950.

(annealing), does not seem to work well, even under high pressure. But replacing oxygen ions is not the only way to introduce holes into high- T_c compounds. For example, because yttrium ions (Y^{3+}) and calcium ions (Ca^{2+}) are almost identical in size, it is possible to introduce holes into YBCO by replacing some of the Y^{3+} by Ca^{2+} .

Now imagine that you could balance the number of holes lost at the grain boundary through oxygen depletion with holes from an appropriate number of Ca^{2+} ions substituted for Y^{3+} in the same region. You might be able to 'repair' the weak link with a low-resistance — and hopefully superconducting — bridge. This is what Hammerl *et al.* have done. Their first attempt⁸ at this approach led to the grains themselves, as well as the boundaries, receiving the same amount of Ca^{2+} , so the former became 'overdoped' (too many holes) and T_c dropped from 91 K to 77 K. Not so impressive if you want to use liquid nitrogen for a coolant, because a superconductor needs to be well below its transition temperature to carry a useful amount of current. In their latest work, Hammerl *et al.*¹ have grown a calcium-doped YBCO film over an undoped one, and found that some of the Ca^{2+} appears to migrate into the grain boundaries, partially 'healing' them. This improves superconducting current flow between grains three-

to sixfold, without degrading current flow within grains or adversely reducing T_c . The authors suggest that their result could have profound consequences for future developments of high- T_c technology. I agree.

Present high- T_c wires — or more precisely, tapes — are based on lead-stabilized $Bi_2Sr_2Ca_2Cu_3O_x$ (BSCCO) drawn into a series of long, fine filaments enclosed in silver⁹. It is now possible to produce kilometre lengths of such tape that carry supercurrents in excess of 100 amperes (A) at 77 K. The main barrier to wider use of this 'first generation' technology is cost (principally of the silver). The price-to-performance ratio of superconducting wire is measured in US dollars per unit of supercurrent per unit of length. For example, the performance of Nb_3Sn , used in many high-field laboratory research magnets, is about $\$5-6 \text{ kA}^{-1} \text{ m}^{-1}$, and $NbTi$, the mainstay of MRI and particle-collider magnets, is roughly $\$0.90 \text{ kA}^{-1} \text{ m}^{-1}$. For comparison, several companies offer BSCCO tape at $\$300 \text{ kA}^{-1} \text{ m}^{-1}$, with estimates that this figure may drop to $\$50 \text{ kA}^{-1} \text{ m}^{-1}$ with mass production. But this is still too high. Fortunately, a second-generation technology is being developed, and the findings of Hammerl *et al.* may mean that much cheaper alternatives are not far off.

This second-generation technology is based on 'coated conductors' — films of

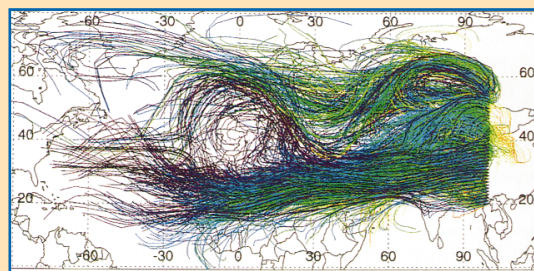
Atmospheric physics

Sky-high Asian imports

The Tragedy of the Commons has long been seen to apply to air pollution. Many nations freely dispose of their emissions, but others — downwind, and maybe far away — will be most affected by the consequences.

In this context, the paths of emissions from Asia across the Pacific Ocean, and their potential effects on air quality in the United States, have been well investigated. Reginald Newell and Mathew Evans of the Massachusetts Institute of Technology have taken a different perspective, as they report in *Geophysical Research Letters* (16, 2509–2512; 2000). They have investigated the question of how much pollution reaches Asia from sources elsewhere.

Newell and Evans set up particles along an imaginary wall, extending meridionally from north of Bangkok to



Siberia. Data on wind strengths and directions allowed them to reconstruct the particles' possible routes through the atmosphere. The picture here shows the resulting 'spaghetti plot', with each line representing a path that would have taken an air parcel, starting at an altitude of 2.5 km or more, to the imaginary wall within five days.

A detailed analysis of the complete data set reveals that in January and February, 30–40% of the air that arrives at longitude 100° E has passed

over Europe. Considering European emission levels, and the tendency of winter storms to swirl pollution up through the atmosphere, European air up to a height of about 10 km is likely to be far from pristine. So, in winter, Europe may contribute significantly to Asian pollution.

So much for the dynamics. At this point, the researchers hand over to the atmospheric chemists, who will have to measure the composition of the atmosphere above Asia to confirm or disprove the point.

Heike Langenberg

Neurobiology

Bundling up excitement

Jeffrey D. Rothstein

YBCO deposited on specially prepared metal tapes¹⁰. Such films have grain boundaries at angles of 5–10° within the CuO₂ planes and supercurrent densities (J_c) bordering on 10⁶ A cm⁻² over metre lengths. Even though the average grain boundary angles are relatively small, the main impediment to making longer high-quality wires (even a few metres) remains the low J_c between grains. Calcium doping may be the answer.

It has been estimated that if the performance of coated conductors at the metre-level could be scaled to kilometre lengths, the basic manufacturing cost of coated-conductor tape would approach \$1 kA⁻¹m⁻¹. Successful commercialization of coated conductors is considered so essential to future electricity infrastructure and use that major government-sponsored efforts have been launched in Japan and the United States to assist private industry in its development.

Much remains to be understood about what really happens within calcium-doped grain boundary junctions in YBCO. How does the calcium get there? Are there other, and possibly simpler and more efficient, methods of inserting Ca²⁺ into the boundary? There is also the issue of the relationship of the grain-boundary J_c as measured by Hammerl *et al.*, to the macroscopic J_c relevant to wire performance, especially the degree to which the 'weak link' nature of the boundary — that is, its sensitivity to external magnetic fields — remains unchanged despite the improvement of J_c in an internal field. A forthcoming paper¹¹ will report a 30% improvement in J_c between grains in high magnetic fields as a result of Ca²⁺ doping of a bicrystal boundary (at a rotation angle of 5° commonly found in coated conductors). Although these data were measured at 44 K because the entire sample was overdoped with holes, thus reducing overall T_c , the same improvement might be expected in samples at 77 K in which only the boundary region is doped.

The use of preferential calcium doping of YBCO grain boundaries to improve the properties of coated conductors, as supported by the data of Hammerl *et al.*, is the best idea that's come along in the five years since coated conductors first emerged. As they say in the adverts, a diet high in calcium is good for you — and maybe for high- T_c wire as well. ■

Paul M. Grant is at the Electric Power Research Institute, 3412 Hillview Avenue, Palo Alto, California 94303, USA.
e-mail: pgrant@epri.com

- Hammerl, G. *et al.* *Nature* **407**, 162–164 (2000).
- Kirtley, J. *et al.* *Phys. Rev. Lett.* **76**, 1336 (1996).
- Dimos, D. *et al.* *Phys. Rev. Lett.* **61**, 218 (1988).
- Stolbov, S. V. *et al.* *Phys. Rev. B* (in the press).
- Stolbov, S. V. *et al.* *Superconduct. Sci. Tech.* **12**, 1071 (1999).
- Babcock, S. & Larbalestier, D. *Physica C* **277**, 183 (1994).
- Gurevich, A. & Pashitskii, E. A. *Phys. Rev. B* **57**, 13878 (1998).
- Schmehl, A. *et al.* *Europhys. Lett.* **47**, 110 (1999).
- Vase, P. *et al.* *Superconduct. Sci. Tech.* **13**, R71 (2000).
- Grant, P. M. *Nature* **375**, 107–108 (1995).
- Daniels, G. A. *et al.* *Appl. Phys. Lett.* (in the press).

Glutamate is a ubiquitous amino acid; every cell in our body is packed full of it. But in the brain it functions in a unique way — as a neurotransmitter. In nerve cells, glutamate is bundled into tiny, membrane-encased spheres called synaptic vesicles. When a neuron is stimulated, these vesicles rush to its outer membrane, releasing glutamate outside the cell. Glutamate then rapidly excites other neurons. Neurons are usually defined by the neurotransmitter they release; most release glutamate (they are 'glutamatergic'). As such, glutamate forms the basis for most of the brain's activity. It is responsible for your ability to see this journal, and to read and understand the words on this page. Over the past decade we have learned much about how glutamate excites neurons, but we know little about what gives a neuron the ability to release glutamate. On page 189 of this issue¹ Takamori and colleagues provide the answer, in the shape of the protein responsible for packing glutamate into synaptic vesicles.

A neurotransmitter itself is defined in part by the molecular mechanisms that make it, release it and mop it up from the extracellular environment after it has done its job. Over the past 30 years we have come a long way towards understanding how glutamate is made and cleared up. Now it is the turn of the release mechanisms to face the glare of the researchers' spotlight.

When packets of glutamate are released from a stimulated neuron (Fig. 1), the neurotransmitter diffuses across the 'synaptic cleft' to the receiving neuron. There, it activates a series of receptors, leading to stimulation of that neuron, too. The glutamate then needs to be cleared up, and transmembrane glutamate transporters found on the surface of nearby astroglial (supporting) cells are the major synaptic vacuum cleaner. Finally, within the signalling (presynaptic) neuron, more glutamate has to be bundled into synaptic vesicles, ready for another round of stimulation.

It turns out that most of the new glutamate is synthesized from another amino acid, glutamine, which is released from astroglial cells and taken up by the presynaptic neuron. An enzyme called phosphate-activated glutaminase converts the glutamine to glutamate, which is then bundled into synaptic vesicles. Biochemical studies indicated that a unique transporter was involved in the repackaging, but, until now, the molecular nature of this transporter has eluded researchers. In fact, the repackaging is one of the critical steps in the overall process of glutamate release,

so the expression of the molecule required for this process might be what makes a neuron able to release glutamate.

Two years ago, it was discovered that the brain-specific, sodium-dependent phosphate transporter (BNPI) is found mainly in the presynaptic terminals of glutamatergic neurons². Moreover, electron microscopy showed that BNPI localizes specifically to the synaptic vesicles in those neurons². But it was not clear whether BNPI has a function in neurotransmitter physiology. Intriguingly, though, a protein called EAT-4 — a relative of BNPI in the nematode worm *Caenorhabditis elegans* — has an essential role in glutamate transmission³. So it was suspected that BNPI might have a similar function in mammals.

In an elegant series of studies, Takamori

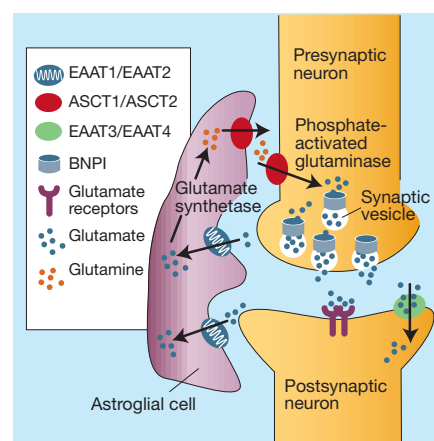


Figure 1 What gives a glutamate-releasing neuron its identity? Most of the neurons in the mammalian brain release glutamate as a neurotransmitter. When the presynaptic neuron is stimulated, synaptic vesicles containing glutamate merge with the neuron's plasma membrane and release their contents outside the cell. Glutamate diffuses to the postsynaptic neuron and binds to receptors there, activating the postsynaptic cell. The released glutamate then needs to be cleared away. This is achieved mainly by the glutamate transporters EAAT1 and EAAT2 (found on astroglial — support — cells), and to a lesser extent by transporters on the postsynaptic neuron (EAAT3 and EAAT4). In astroglial cells, glutamate is enzymatically degraded to glutamine. Glutamine may then be supplied to neurons, through the neutral-amino-acid transporters ASCT1 and ASCT2. Once in the neuron, glutamine is converted to glutamate. Takamori *et al.*¹ and Bellocchio *et al.*⁴ show that glutamate is then repackaged into vesicles by a protein called BNPI. Expression of BNPI in neurons that do not normally release glutamate is sufficient to make them do so¹.

*et al.*¹, along with Bellocchio *et al.*⁴, now show that the expression of BNPI in cell lines leads to specific, ATP-dependent uptake of glutamate into vesicles. Does this mean that this protein is what makes a neuron glutamatergic? Takamori *et al.*¹ go on to show that the expression of BNPI in a cell line that can be stimulated by the excitatory neurotransmitter acetylcholine results in the release of packets of glutamate from the cells. Finally, they convincingly show that neurons that normally release the neurotransmitter GABA can be persuaded to release glutamate instead by the artificial expression of BNPI.

So it seems that the vesicular glutamate transporter BNPI may be the best marker by which to define a glutamatergic neuron. But in some ways, all that these groups^{1,4} have done is identify another vesicle-bound neurotransmitter transporter. What makes the work much more exciting is that we may now have a way of specifically controlling neurotransmission within glutamatergic neuronal circuits. A wide range of neurological diseases are characterized by the aberrant regulation of glutamate. These include acute stroke (brain injury resulting from a sudden loss of blood or oxygen to the brain), as well as more chronic neurodegenerative disorders such as Huntington's disease and amyotrophic lateral sclerosis. Both vesicular⁵ and transmembrane⁶ glutamate transporters have been implicated in these diseases.

The past decade has seen an enormous effort by pharmaceutical firms to develop potent 'anti-glutamate' agents as neuroprotective compounds. But many of these have been either ineffective or too toxic. The new results afford us the opportunity to control the release of glutamate by manipulating its loading into vesicles. But some questions remain unanswered. For example, BNPI was identified as a phosphate transporter. Can it function as such *in vivo* and, if so, which is its main role — to uptake glutamate or to transport phosphate? In addition, BNPI is not found in all of the known neuronal pathways along which glutamate travels, so other types of vesicular glutamate transporter may exist. If so, we may one day be able to develop drugs that are specific to these different types, allowing us precise control of the glutamatergic neuronal circuitry.

Jeffrey D. Rothstein is in the Department of Neurology, Johns Hopkins University, Meyer 6-109, 600 North Wolfe Street, Baltimore, Maryland 21287, USA.

e-mail: jrothste@welchlink.welch.jhu.edu

1. Takamori, S., Rhee, J. S., Rosenmund, C. & Jahn, R. *Nature* **407**, 189–194 (2000).
2. Bellocchio, E. E. *et al.* *J. Neurosci.* **18**, 8648–8659 (1998).
3. Lee, R. Y., Sawin, E. R., Chalfie, M., Horvitz, H. R. & Avery, L. *J. Neurosci.* **19**, 159–167 (1999).
4. Bellocchio, E. E., Reimer, R. J., Freneau, R. T. Jr & Edwards, R. H. *Science* **289**, 957–960 (2000).
5. Li, H., Li, S. H., Johnston, H., Shelbourne, P. F. & Li, X. J. *Nature Genet.* **25**, 385–389 (2000).
6. Lin, C. G. *et al.* *Neuron* **20**, 589–602 (1998).

Global change

Plankton cooled a greenhouse

Birger Schmitz

Scientists who can perform laboratory experiments are lucky — a megalomaniac climatologist can only dream of putting an Earth-like planet in a giant test tube, pumping billions of tonnes of CO₂ into its atmosphere, and registering the effects on life and climate. Fortunately, there are other approaches. At the Palaeocene/Eocene (P/E) boundary 55 million years ago, nature appears to have done the greenhouse experiment for us. Bains *et al.*¹ (page 171 of this issue) now report that they have identified a rather unexpected response of the oceanic biosphere to dramatically high concentrations of atmospheric CO₂, and temperatures, at this boundary — one that can account for a subsequent reduction in atmospheric CO₂ and cooling.

In the early Palaeogene, 65 to 41 million years ago — a period that includes the Palaeocene and the first half of the Eocene — the Earth was generally much warmer than today. The polar regions were free of continental ice sheets, alligators and turtles thrived on Ellesmere Island at 75° N, and palm trees grew as far north as Kamchatka. For a period of about 150,000 years, 'super-warm' conditions developed at what is known as the P/E thermal maximum. Oxygen isotopic (¹⁸O/¹⁶O) analyses of shells from

marine microorganisms called foraminifera show that surface-water temperatures off the coast of Antarctica rose from about 13 °C to 20 °C (ref. 2). Subtropical regions also became warmer; but the higher the latitude, the greater was the effect^{3,4}. The warming coincides with some of the most dramatic biotic changes since the mass extinctions at the Cretaceous/Tertiary boundary, 65 million years ago. Deep-sea, bottom-dwelling foraminifera suffered major extinctions, while terrestrial mammals and different oceanic plankton groups underwent considerable diversification.

Events at the P/E boundary coincide with a large decline in the ¹³C/¹²C ratio of the carbon dissolved in the ocean^{2–4}. The decline happened rapidly, within a few tens of thousands of years. It is recorded in the shells of both planktonic and deep-sea, bottom-living foraminifera, as well as in the teeth of land mammals, indicating that the entire ocean-atmosphere carbon reservoir altered in composition.

How could such a big change have happened so quickly? Several years ago, Dickens *et al.*⁵ provided a plausible explanation. Under normal temperature conditions, enormous amounts of carbon are stored in ocean sediments as gas hydrates — solid

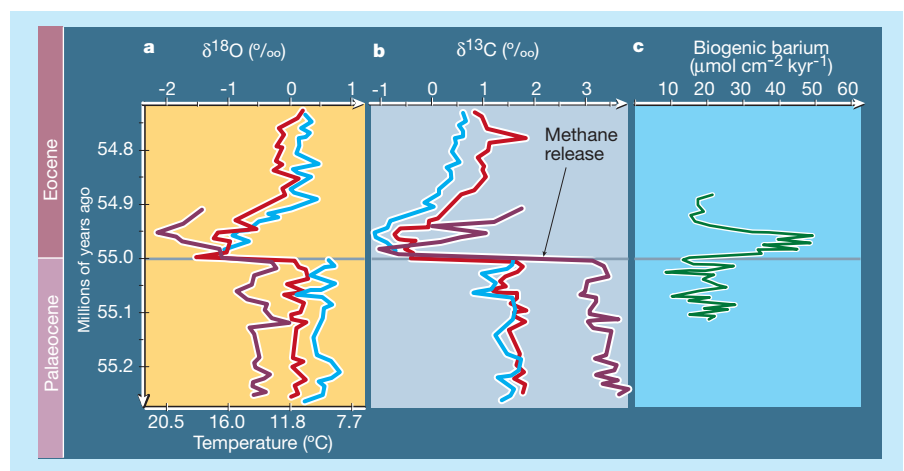


Figure 1 Events at the Palaeocene/Eocene (P/E) boundary, 55 million years ago, inferred from data from an ocean-sediment core drilled in the Weddell Sea, off Antarctica. a, b, Changes in the ¹⁸O/¹⁶O and ¹³C/¹²C isotopic composition of foraminiferal shells² (given as δ¹⁸O and δ¹³C values in parts per thousand), which provide information about temperature and the source of the carbon, respectively. The implication of the δ¹³C record is that there was a massive methane release at the P/E boundary. The isotope data come from foraminifera that lived at three different depths, and show that large changes in temperature and carbon-isotope composition occurred throughout the water column — *Nuttallides truempyi* (blue) lived on the sea floor, at a depth of about 2,000 m; *Acarinina praepentacamerata* (purple) and *Subbotina* spp. (red) were both planktonic, but shallow- and deep-dwelling respectively. c, As described by Bains *et al.*¹, the biogenic barite record implies that productivity in surface waters increased substantially at and beyond the P/E boundary. They invoke the drawdown of carbon from the atmosphere to explain the return to cooler conditions.

crystals of water and methane. If the ocean warms, these gas hydrates may become destabilized, releasing methane gas which becomes oxidized to CO_2 . The methane is enriched in the light ^{12}C relative to ^{13}C , showing that it was originally formed by bacteria which fractionate isotopes during their metabolism. Dickens *et al.* calculated that a release of $1,500 \times 10^9$ tonnes of methane could account for the large change in carbon isotope ratios at the P/E boundary. On a millennial scale, this would imply a rate of greenhouse-gas emission comparable to current anthropogenic levels.

So what happens if 1,500 gigatonnes of methane are released to the ocean-atmosphere system? Regional differences in temperature are the main driving force for winds and ocean circulation. Palaeoclimate modellers have therefore generally assumed that, with preferential warming at high latitudes, an 'equable' Earth would develop during the P/E thermal maximum. Oceanic circulation would become more sluggish, fewer nutrients would reach the sunlit surface zone, and plankton productivity and photosynthesis would fall.

But some data point to an opposite conclusion. It seems that certain species of plankton characteristic of high-productivity regions flourished all over the world at the P/E boundary⁶. And in rock sections in the Middle East, formed from sediments that were originally under sea, a large peak in biogenic barite (a mineral of barium sulphate) at the P/E boundary indicates a dramatic increase in biological productivity⁷. Biogenic barite has proved a reliable proxy for surface-water biological productivity in the open oceans of the past^{7,8}.

The paper by Bains *et al.*¹ not only provides new evidence that oceanic productivity did indeed increase, but also provides a feasible mechanism for how an episode of greenhouse warming may end. They show that in two widely separated ocean drilling cores — one from off Antarctica, the other from the western North Atlantic — the distribution of biogenic barite is a mirror image of the $^{13}\text{C}/^{12}\text{C}$ and $^{18}\text{O}/^{16}\text{O}$ curves across the P/E thermal maximum (Fig. 1). This implies that biological productivity increased as methane was released and temperature increased (though there is a chicken-and-egg problem here); and productivity decreased as climatic conditions returned to normal. Bains *et al.* propose that higher productivity and the resulting sequestering of excess carbon in the oceans, through photosynthesis, was the feedback mechanism required to bring levels of atmospheric CO_2 and temperatures back to normal.

Proponents of the Gaia theory, which says that the biosphere regulates climate⁹, will love this interpretation. But carbon is not a productivity-limiting nutrient, and Bains *et al.* say there was probably a secondary

feedback between high levels of CO_2 and productivity, or that the relationship is just coincidental. They speculate that greater humidity in a greenhouse world, and consequent increased runoff of water from land, or volcanic fallout (or both), fertilized the oceans' surface waters.

Can we rely on the barite proxy for productivity? I believe we can. There is firm evidence that barite crystals have accumulated in sediments under high-productivity areas in recent times, and that this signal is retained in sediments as old as the early Palaeogene^{7,8}. Bains *et al.* show that there are unaltered barite crystals in the P/E sediments they analysed.

However, there is still no consensus as to how the crystals form in the ocean. According to one school of thought, the barite precipitates in decaying organic matter settling through the water column; another holds that the crystals form in living organisms. Various single-celled organisms precipitate microscopic crystals of barite in their bodies, possibly as statoliths for orienting themselves in the gravitational field, but no organism has been clearly linked to the abundant crystals under high-productivity regions¹⁰.

There are also other puzzles. The idea of a huge release of methane at the P/E boundary is popular mainly because, so far at least, it is the only realistic explanation

for the observed large and rapid decline in the oceanic $^{13}\text{C}/^{12}\text{C}$ ratio. But two proxy reconstructions of CO_2 concentrations during the P/E thermal maximum have failed to find evidence for substantially increased concentrations^{4,11}. As usual, we need to know more. Is there something wrong with the proxy estimates of atmospheric CO_2 concentrations in the past, or is there another explanation for the decline in $^{13}\text{C}/^{12}\text{C}$ at the P/E boundary? And is ocean productivity increasing in our incipient greenhouse world? The paper by Bains *et al.* will trigger much new work on these questions.

Birger Schmitz is in the Department of Earth Sciences, University of Göteborg, SE-405 30 Göteborg, Sweden.

e-mail: birger@gvc.gu.se

1. Bains, S., Norris, R. D., Corfield, R. M. & Faul, K. L. *Nature* **407**, 171–174 (2000).
2. Kennett, J. P. & Stott, L. D. *Nature* **353**, 225–229 (1991).
3. Thomas, D. J., Bralower, T. J. & Zachos, J. C. *Paleoceanography* **14**, 561–570 (1999).
4. Stott, L. D. *Paleoceanography* **7**, 395–404 (1992).
5. Dickens, G. R., O'Neil, J. R., Rea, D. K. & Owen, R. M. *Paleoceanography* **10**, 965–971 (1995).
6. Crouch, E. M., Bujak, J. P. & Brinkhuis, H. *GFF* **122**, 40–41 (2000).
7. Schmitz, B., Charisi, S. D., Thompson, E. I. & Speijer, R. P. *Terra Nova* **9**, 95–99 (1997).
8. Paytan, A., Kastner, M. & Chavez, F. P. *Science* **274**, 1355–1357 (1996).
9. Lovelock, J. E. *Gaia: A New Look at Life on Earth* (Oxford Univ. Press, 1995).
10. Bertram, M. A. & Cowen, J. P. *J. Mar. Res.* **55**, 577–593 (1997).
11. Sinha, A. & Stott, L. D. *Glob. Planet. Change* **9**, 297–307 (1994).

Cell biology

The specifics of membrane fusion

Suzie J. Scales, Jason B. Bock and Richard H. Scheller

The ability to maintain a diverse set of intracellular compartments, with distinct complements of proteins, is a defining feature of eukaryotic cells. Substances can be transported from one membrane-encased compartment to another, but the compartments maintain their unique identities. Transport occurs in membrane-bounded containers called vesicles, and several protein families have evolved to mediate the budding of a vesicle from the donor compartment, and its transport to and fusion with the target organelle. One of the last steps in the fusion process is overseen by a set of proteins called SNAREs. These have been suggested to be the core machinery that mediates the fusing of two membranes, as well as ensuring that vesicles deliver their cargo to the right compartment^{1,2}. Writing on pages 153, 194 and 198 of this issue^{3–5}, Rothman and colleagues conclude — with some caveats — that SNAREs are indeed important in defining the specificity of vesicle targeting.

SNAREs contain structural features

called α -helices or coils. During membrane fusion, four α -helices from SNAREs found on the vesicle and target membranes come together to form a stable, four-helix bundle or coiled-coil⁶ (Fig. 1a). The formation of SNARE complexes is essential for membrane fusion, so a tremendous amount of research has been dedicated to understanding how these complexes form, and what they do. Rothman and colleagues earlier developed an *in vitro* assay that measures the fusion of liposomes — artificial spheres surrounded by a lipid membrane — reconstituted with neuronal SNARE proteins⁷. This system is ideal for assessing the role of SNAREs in isolation. It has been used to show that, when present on liposomes representing the vesicle and the target, SNAREs — in the absence of other factors — can induce membrane fusion⁷.

As well as being involved in driving membrane fusion, SNARE proteins have been implicated in ensuring the accuracy of vesicle trafficking. There appear to be enough SNAREs, differently localized throughout

intracellular membrane compartments, to confer specificity on the process. For example, a transport vesicle that buds from the endoplasmic reticulum (ER) is destined for the Golgi, but how does the cell make sure that it doesn't fuse with, say, the plasma membrane instead? The idea is that, in this case, the SNAREs on the ER-derived vesicle can form a complex only with those on the Golgi, ensuring specificity (Fig. 1a). Such SNAREs are said to be 'cognate'.

Cognate SNAREs have been characterized for some vesicle-trafficking steps, and several themes have emerged from these studies, the most important of which is the composition of the SNARE complexes. A four-helix bundle always contains one coil from each of four families^{6,8}, abbreviated here for simplicity as A, B, C and D coils (for specialist readers, A designates the Q-SNARE syntaxin, B and C the Q-SNAREs of the SNAP-25 family, and D the R-SNARE of the VAMP family; Fig. 1b).

Rothman and colleagues³⁻⁵ have now used their liposome-based system to investigate whether isolated yeast SNAREs have a role in ensuring fusion specificity. The power of this experimental set-up is that any combination of SNAREs can be put into the 'vesicle' and 'target' liposomes. The authors find that there is a great deal of specificity in terms of which liposomes fuse together. But this specificity comes from different sources.

For example, the physical arrangement (topology) of the SNAREs can confer specificity. The authors tested the SNAREs that (in cells) are involved in transport between the ER and the Golgi, and found that liposome fusion occurred only when the A, B and D coils (from yeast SNARE proteins Sed5, Bos1 and Sec22, respectively) were on one membrane and the C coil (Bet1) was on the other³. No other combination of ER-Golgi SNAREs tested resulted in fusion.

It would be interesting to see whether such topological restrictions are important *in vivo*, and whether they extend to other trafficking steps. They clearly don't apply in the case of the yeast plasma-membrane SNARE (Sec9), because the B and C coils are fused into one protein, preventing the C coil from being on the other membrane. Indeed, the authors show⁴ that fusion involving plasma-membrane SNAREs occurs when A, B and C coils are on the same membrane, with the D coil on the other. This arrangement also works for the vacuolar SNAREs tested^{4,5}. One explanation for the unexpected ER-to-Golgi arrangement could be that Sec22 functions in transport from the Golgi to the ER, whereas the other three SNAREs function in transport from the ER to the Golgi^{9,10}. So Sec22 may not contribute the correct D coil in this case³.

Specificity may also come from the interactions of cognate SNAREs. An intuitively

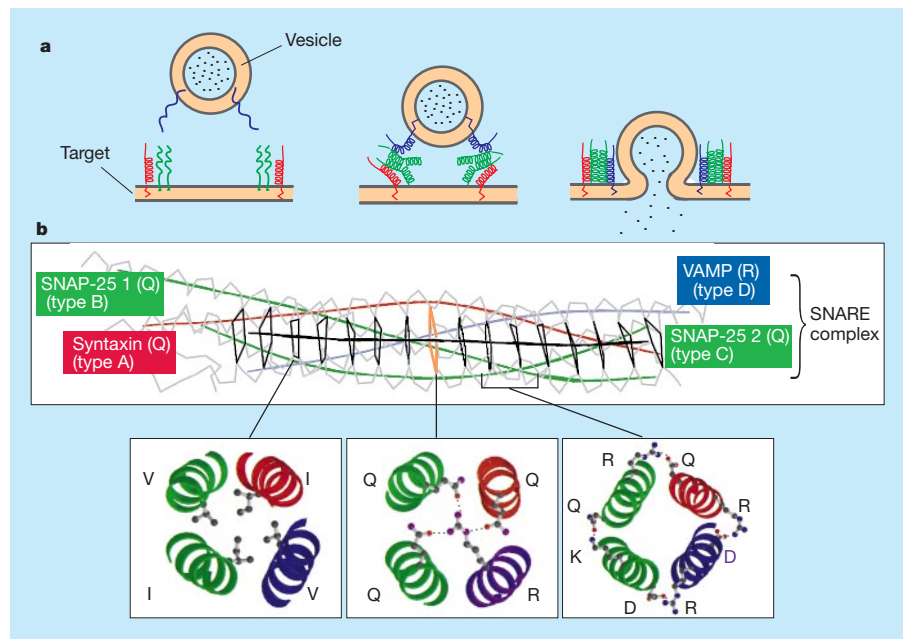


Figure 1 SNARE proteins and vesicle targeting. SNAREs are found on vesicle and target membranes, and between them contribute four 'coil' structures to a SNARE complex. **a**, Rothman and colleagues³⁻⁵ have constructed artificial membrane-encased spheres (liposomes) with different combinations of SNAREs, to assess the role of SNAREs in fusion specificity. They find, as shown previously⁶, that — for vesicle fusion with the plasma membrane — three Q-SNAREs (red, coil type A; green, coil types B and C) are required on the target and one R-SNARE (blue; coil type D) on the vesicle^{4,5}. SNAREs confer some specificity to vesicle fusion³⁻⁵. **b**, Q-SNAREs and R-SNAREs are characterized by a glutamine (Q) or arginine (R) residue, respectively, in the central layer of the SNARE complex. Top, the SNARE complex required for fusion of a synaptic vesicle with the plasma membrane (modified from ref. 6); bottom, the underlying molecular interactions. Black, the hydrophobic layers that mediate the core interactions; orange, the Q/R layer. Details of the interactions are shown underneath. V, valine; I, isoleucine; Q, glutamine; R, arginine; K, lysine; D, aspartic acid. The ball-and-stick structures represent the indicated amino acids; the dotted lines represent hydrogen bonds or salt bridges that stabilize interactions between SNAREs.

appealing hypothesis was that only cognate SNAREs could form complexes. But this theory was called into question by *in vitro* studies showing that, although an A coil, a B coil, a C coil and a D coil were strictly required for complex formation, it did not matter from which SNARE protein these coils originated^{8,11}. However, another study¹² showed that, in a particular cell line, only cognate SNAREs, when added in solution, could compete with membrane-bound SNAREs and inhibit vesicle fusion — with a couple of exceptions, non-cognate SNAREs in solution could not. Rothman and colleagues⁴ again used the liposome assay to look at this problem. They found that one of each coil — A, B, C and D — are absolutely required. For many incorrect SNARE combinations (such as an A coil, two Bs and a D; or an A, a B and two Ds), fusion does not occur.

But the true test of specificity is to find out whether (for example) different D coils can substitute for each other, rather than for C coils. The authors find several examples in which cognate A, B, C and D coils (for example Vam3, Vam7, Vti1 and Nyv1) produce more fusion than non-cognate coils (Vam3, Vam7, Vti1 and either Snc1 or Sec22). They did not investigate what

happens when A coils are swapped, leaving open the possibility that some non-cognate SNARE interactions might still result in fusion. Indeed, there are examples in which non-cognate A, B, C and D coils result in fusion. With the plasma-membrane A, B and C coils, fusion can occur when any D coil is used⁴. So isolated SNAREs cannot solely account for the specificity observed in vesicle trafficking.

As the authors say⁴, it is likely that specificity is not uniquely encoded in the inherent ability of SNAREs to form complexes with each other. Given that SNAREs bind to each other through their conserved features (Fig. 1b), this is perhaps not surprising. Other proteins required earlier in the fusion process — such as members of the Sec1 and Rab families — may help cognate SNAREs to recognize each other. And the organization of the cell probably adds to specificity. For example, an ER-derived vesicle is more likely to encounter the Golgi than the plasma membrane, so fusion with the plasma membrane is unlikely to occur, even if the ER and plasma-membrane SNAREs were able to form a complex *in vitro*. It seems that the accuracy of vesicle targeting is safeguarded not through a single lock-and-key inter-

action between SNAREs, but rather through several layers of constraints — a situation common to many biological processes. ■

Suzie J. Scales, Jason B. Bock and Richard H. Scheller are at the Howard Hughes Medical Institute, Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305-5345, USA.
e-mail: sjscales@cmgm.stanford.edu

1. Bennett, M. K. & Scheller, R. H. *Proc. Natl Acad. Sci. USA* **90**, 2559–2563 (1993).

2. Söllner, T. *et al. Nature* **362**, 318–324 (1993).
3. Parlati, F. *et al. Nature* **407**, 194–198 (2000).
4. McNew, J. A. *et al. Nature* **407**, 153–159 (2000).
5. Fukuda, R. *et al. Nature* **407**, 198–202 (2000).
6. Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. *Nature* **395**, 347–353 (1998).
7. Weber, T. *et al. Cell* **92**, 759–772 (1998).
8. Yang, B. *et al. J. Biol. Chem.* **274**, 5649–5653 (1999).
9. Spang, A. & Schekman, R. *J. Cell Biol.* **143**, 589–599 (1998).
10. Cao, X. & Barlowe, C. *J. Cell Biol.* **149**, 55–66 (2000).
11. Fasshauer, D., Antonin, W., Margittai, M., Pabst, S. & Jahn, R. *J. Biol. Chem.* **274**, 15440–15446 (1999).
12. Scales, S. J. *et al. Neuron* **26**, 457–464 (2000).

X-ray astronomy

Imaging black holes

Nicholas White

Black holes hold an almost mythical attraction for layperson and scientist alike. A black hole is an object so massive and compact that gravity prevents even light from escaping. The gravitational effect of a black hole on nearby objects provides compelling indirect evidence that they exist, but the ultimate proof has yet to come — a direct image of the ‘black dot’.

On page 160 of this issue, Cash *et al.*¹ present the first laboratory demonstration of an X-ray interferometer that will be useful to astronomers. Their approach will make it much easier to achieve the angular resolution of 0.1 to 1.0 microarcseconds required to obtain an X-ray image of black holes in the centre of nearby galaxies. Astronomers divide up the sky into angular degrees, so that 90° is the distance from the horizon to a point directly overhead (there are 60 arcminutes in a degree and 60 arcseconds in a minute). Apart from satisfying our curiosity as to what the region surrounding a black hole looks like, this advance will allow us to directly observe effects predicted by Einstein’s theory of general relativity under the most extreme gravity fields known. It will also provide a formidable tool that will open new vistas on a wide range of astronomical phenomena.

The X-ray band is the prime hunting ground for finding and studying black holes, as shown by the first bona fide ‘black hole candidate’, Cygnus X-1, an X-ray source discovered in the 1960s. Bright X-rays are the result of large amounts of gravitational energy being released as the black hole attracts material from a nearby star or within its host galaxy. This material forms a swirling, orbiting disk falling towards the black hole — much like the flow of water down a drain (Fig. 1). Close to the ‘event horizon’, the theoretical border of a black hole inside which nothing can escape, friction superheats the material to many millions of kelvin, which is mostly radiated as X-rays. The black hole’s strong gravity

causes distortions of space-time that are imprinted on the emerging X-rays. Observations of supermassive black holes in the centre of nearby galaxies have already revealed this signature in the spectral features of X-rays².

Increasing the angular resolution of telescopes is one of astronomy’s main goals, but it is never easy. Even the most perfectly shaped telescope is ultimately limited by the size of its aperture, also known as the diffraction limit. This is dictated by the wavelength of the incoming light divided by the diameter of the telescope. The bigger the telescope, the better the angular resolution it can achieve. The Hubble Space Telescope has a 2.4-metre diameter with an angular resolution of 0.1 arcsecond, which is close to the diffraction limit. Achieving microarcsecond resolution would require a 100,000-fold increase in the Hubble telescope diameter to 240 kilometres.

Fortunately, there is a way of achieving such resolution without building impossibly large telescopes. An interferometer combines the light from several small telescopes to create an image with a resolution as if it had come from a much larger telescope. The light waves from each telescope interfere with each other to create interference fringes (bands of low and high intensity), which can be transformed back into real images inside a computer. For interferometers at most wavelengths, the distance between the telescopes replaces the diameter of the telescope in determining the diffraction limit. Radio astronomers first used this technique to make huge gains in angular resolution with telescope separations spanning continents and even out into space.

X-ray telescopes in general are difficult to build because X-rays reflect only at a very shallow angle to the optical surface (1 degree or less), referred to as the grazing incidence. To obtain a true focus, they must be reflected twice from precisely constructed hyperbolic and parabolic surfaces. These surfaces are, in effect, nested cylinders that are expensive to

shape to the required precision. Complicating matters further, X-ray telescopes must be placed in space, because X-rays do not penetrate the Earth’s atmosphere. The recently launched Chandra X-ray Observatory is the state of the art in X-ray imaging, whose optics alone cost several hundred million US dollars to build. Chandra achieves an impressive resolution of about 0.5 arcseconds, yet it is still far from the diffraction limit. Building a diffraction-limited X-ray telescope, let alone an X-ray interferometer capable of imaging the cauldron surrounding a black hole, has always seemed a distant dream.

Cash *et al.*¹ take what at first seems to be a disadvantage — that X-rays reflect only at shallow angles — and turn it into an advantage. Instead of using expensive, precisely figured optics to focus the X-rays, they

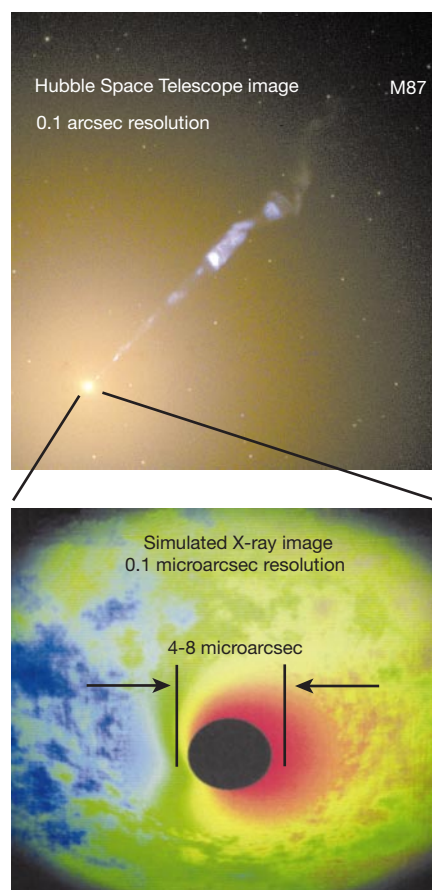


Figure 1 Imaging a black hole requires an enormous improvement in telescope power — at least a million times better resolution than the Hubble Space Telescope, or the recently launched Chandra X-ray Observatory. Top, a Hubble Telescope image of the M87 galaxy core, where a black hole with a mass three billion times that of the Sun is likely to reside. Bottom, a simulation of what the black hole might look like³ if you were looking down on a disk of material swirling around the hole. The angular size that the black-hole event horizon subtends on the sky is between 3 and 6 microarcseconds, depending on whether or not the black hole is maximally rotating.

instead use two sets of more easily made, flat mirrors to steer incoming X-ray beams together to create interference fringes. A two-dimensional image is created by combining many sets of these fringes taken at different rotation angles. Because the X-rays are reflected at shallow angles, the permitted variations in the positioning of the flat mirrors are about 100 times greater than for a traditional (normal incidence) mirror operating at the same wavelength. If you placed an X-ray detector 500 kilometres behind the mirrors, the fringes would be amplified by the distance, and so could be measured with detectors that exist today.

There are still some technological hurdles, however. Even at very short X-ray wavelengths, a telescope separation of 100–1,000 metres is needed to achieve the required angular resolution. This would require a fleet of up to 33 spacecraft carrying optical mirrors, flying in formation with a spatial precision of 20 nanometres, plus a detector spacecraft 500 kilometres behind the mirror. This is daunting by today's standards, but probably no more so than missions under consideration by the US and European space agencies (NASA and ESA), such as the Darwin infrared space interferometer, which could search for Earth-size planets outside our Solar System. A 'pathfinder' mission to build an X-ray interferometer with a one-metre separation between the telescopes, so that the X-ray optics are all on one spacecraft, is a reasonable first step. This is already under study at NASA, using Cash and colleagues' technique as the starting design (Fig. 2). The Pathfinder would be a precursor to a much larger Microarcsecond X-ray Imaging Mission (MAXIM) required to image a black hole. But even as a first step, the Pathfinder would provide an impressive 1,000-fold

improvement over the Chandra X-ray Observatory, allowing astronomers to study the coronae of other stars.

Nicholas White is at the Laboratory for High Energy Astrophysics, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

e-mail: nwhite@milkyway.gsfc.nasa.gov

1. Cash, W., Shipley, A., Osterman, S. & Joy, M. *Nature* **407**, 160–162 (2000).
2. Tanaka, Y. *et al.* *Nature* **375**, 659–661 (1995).
3. Bromley, B. C., Miller, W. A. & Pariev, V. I. *Nature* **391**, 54–56 (1998).

http://chandra.nasa.gov

Signal transduction

Lipid rafts and insulin action

Michael P. Czech

One of the greatest medical breakthroughs of the twentieth century was the demonstration in 1922 of insulin's ability to alleviate some of the symptoms of diabetes, including high levels of glucose in the blood. Insulin has received special attention ever since: the amino-acid sequence of the insulin protein was the first to be decoded, and the insulin gene was one of the first to be cloned. Tremendous effort has also gone into understanding how insulin works. We know that it stimulates muscle and fat cells to remove glucose from the blood, by encouraging an intracellular glucose transporter, called GLUT4, to move to the plasma membrane of these cells. But we know few of the details of how insulin does this, beyond the fact that a particular cellular signalling pathway involving lipids is necessary, but may not on its own be sufficient¹. On page 202 of this issue, Baumann and colleagues² offer a possible solution: they describe a second insulin-stimulated signalling pathway in human cells that may also be required to stimulate glucose uptake.

Insulin signalling to muscle and fat cells starts with the binding of insulin to its receptor on the cell surface. The receptor is an enzyme that phosphorylates (adds phosphate groups to) tyrosine residues in other proteins. Like many other such 'receptor tyrosine kinases', the insulin receptor activates several divergent signalling pathways to regulate different cellular processes. One such pathway is represented by the p85/p110-type phosphatidylinositol-3-OH kinase (PI(3)K for short), which phosphorylates the 3' position of the inositol head group of phosphatidylinositol lipids. This results in molecules such as phosphatidylinositol-3,4,5-trisphosphate (PIP₃), which signals through unknown proteins to GLUT4 (Fig. 1, overleaf).

If this pathway is blocked, so too is the insulin-dependent regulation of GLUT4 (refs 3, 4). And adipocytes (fat cells) from mice lacking insulin-receptor substrate (IRS)-1, a linking or 'adaptor' protein that is phosphorylated by the insulin receptor and which recruits PI(3)K, show the expected drop in insulin-mediated glucose uptake³. These and many other findings converged a

few years ago to suggest that figuring out how insulin signals to GLUT4 was simply a matter of identifying the downstream elements in this PI(3)K pathway.

Alas, this simplicity has been eroded by recent events. For example, the insulin-mediated formation of IRS-PI(3)K complexes can be greatly reduced by various experimental manipulations, without affecting the stimulation of glucose uptake⁶. The stimulation of integrin receptors in adipocytes mimics insulin in causing recruitment of PI(3)K to insulin-receptor substrates and activation of a downstream effector, protein kinase B/AKT, without affecting GLUT4 (ref. 7). And a membrane-permeant analogue of PIP₃ also failed to stimulate glucose transport unless insulin was present⁸. All this could be explained if only a very small amount of PIP₃ is required for insulin action, like the spark in an engine, and if another (unknown) pathway, analogous to the engine's motor, has the major role in driving the movement of GLUT4 to the membrane. This concept started a race to find other possible signalling pathways.

The pathway suggested by Baumann *et al.*² involves the recruitment of another adaptor protein, CAP, to the insulin-bound insulin receptor. CAP in turn recruits a protein called Cbl, and also interacts with the flotillin protein. When Cbl is phosphorylated by the insulin receptor, flotillin directs the Cbl-CAP complex to invaginations in the plasma membrane (caveolae), or to regions of the plasma membrane with a high concentration of lipids (lipid rafts) (Fig. 1). This localization is likely to be needed for glucose uptake: expression of truncated, non-functional CAP in cultured adipocytes blocked the binding of Cbl to lipid rafts, the insulin-stimulated movement of GLUT4 and the uptake of glucose. What it didn't block were signalling events immediately downstream of PI(3)K.

Might this insulin-triggered signalling pathway be the motor to the PI(3)K pathway's spark? In its favour is the fact that the expression of CAP is high in muscle and fat, and is upregulated in adipocytes in response to thiazolidinediones — drugs that enhance cellular sensitivity to insulin. One of the

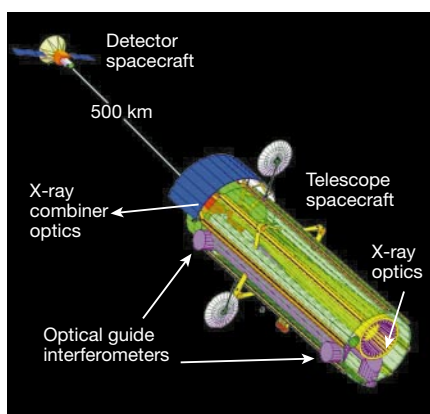


Figure 2 A Pathfinder mission to demonstrate X-ray interferometry in space. This is how the instrument developed by Cash *et al.*¹ in the lab would be scaled up to make a practical space telescope. A mission is already under consideration at NASA for a launch in around 2015 and will provide angular resolution 1,000 times greater than the recently launched Chandra X-ray Observatory.

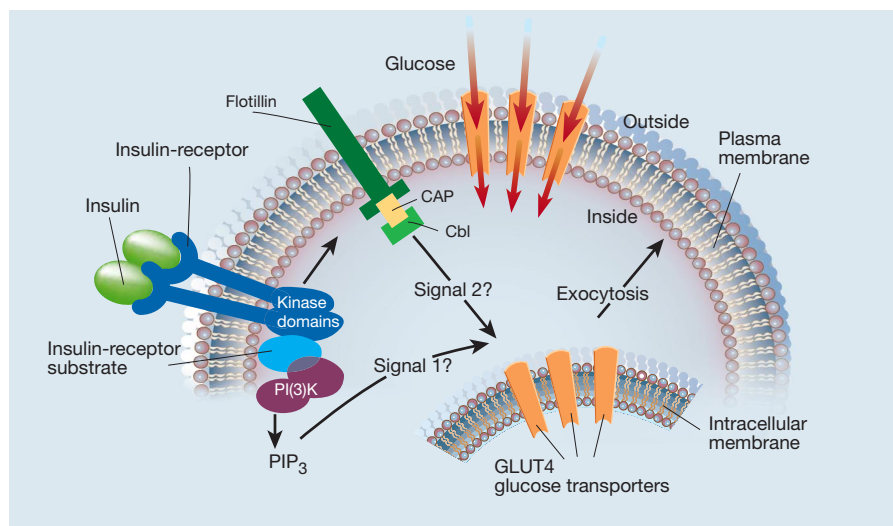


Figure 1 Pathways for glucose uptake in muscle and fat cells. When insulin binds to its receptor on the cell surface, intracellular membranes containing the glucose transporter GLUT4 are induced to fuse with the plasma membrane (exocytosis), allowing the cell to take up glucose. There may be two pathways by which the insulin signal is transmitted to the GLUT4-bearing membranes inside the cell. Both pathways start with the phosphorylation of key substrates by the activated insulin receptor. One pathway, signal 1, involves the phosphorylation of insulin-receptor substrates and then the recruitment to them of phosphatidylinositol-3-OH kinase (PI(3)K). PI(3)K generates phosphatidylinositol-3,4,5-trisphosphate (PIP₃), a membrane lipid, which signals through unknown downstream effectors to regulate GLUT4. The results of Baumann *et al.*² indicate that the second pathway (signal 2) may involve the recruitment of flotillin–CAP–Cbl complexes to particular regions in the plasma membrane.

mysteries of insulin signalling to GLUT4 is the molecular basis of its remarkable tissue specificity, as it occurs in only muscle and fat. Even when GLUT4 and insulin receptors are artificially placed in other cell types, insulin does not stimulate GLUT4 movement to the plasma membrane. This phenomenon, and the fact that Cbl is not tyrosine-phosphorylated in these other cell types, could be explained by the lack of CAP in these cells.

The connection of the flotillin–Cbl–CAP complexes to cholesterol-rich caveolae or lipid raft domains in the plasma membrane of cultured adipocytes is also intriguing. Insulin receptors may be associated with caveolae, a connection that might enhance insulin signalling⁹. And one of the substrates for PI(3)K, phosphatidylinositol-4,5-bisphosphate (from which PIP₃ is generated), is apparently concentrated in cholesterol-enriched domains of the plasma membrane. Although preliminary and fragmentary at present, these observations might reflect the coordinated localization of several signalling pathways involved in GLUT4 regulation.

But, like many odd discoveries, these results raise questions. For example, analysis of human Cbl and its nematode and fruitfly counterparts indicates that its main function in other cell types may be to dampen signalling from receptor tyrosine kinases. It does so by promoting a pathway by which the receptors are degraded after being activated by ligands¹⁰. So why doesn't Cbl also restrain the insulin-receptor tyrosine kinase, rather than seeming to promote its signalling? Does

the truncated Cbl protein modulate GLUT4 transport, too? There are two other forms of mammalian Cbl; can they compensate for each other?

Perhaps the most important question is whether the complex discovered by Baumann *et al.* actually signals in response to insulin (Fig. 1), or whether it simply provides a constitutive, background activity required in some other way for GLUT4 regulation. It is still possible that the PI(3)K signalling pathway is the only one that needs to be activated by insulin. Either way, there are a lot of molecular details to be worked out. For example, does CAP function through signalling proteins known to be recruited to Cbl, or through proteins that have not yet even been discovered? If the history of studying insulin is any indication, the answers will provide us with yet more questions.

Michael P. Czech is in the Program in Molecular Medicine and the Department of Biochemistry and Molecular Biology, 373 Plantation Street, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.

e-mail: michael.czech@umassmed.edu

1. Czech, M. P. & Corvera, S. *J. Biol. Chem.* **274**, 1865–1868 (1999).
2. Baumann, C. A. *et al. Nature* **407**, 202–207 (2000).
3. Cheatham, B. *et al. Mol. Cell. Biol.* **14**, 4902–4911 (1994).
4. Sharma, P. M. *et al. J. Biol. Chem.* **273**, 18528–18537 (1998).
5. Araki, E. *et al. Nature* **372**, 186–190 (1994).
6. Sharma, P. M., Egawa, K., Gustafson, T. A., Martin, J. L. & Olefsky, J. M. *Mol. Cell. Biol.* **17**, 7386–7397 (1997).
7. Guilherme, A. & Czech, M. P. *J. Biol. Chem.* **273**, 33119–33122 (1998).
8. Jiang, T. *et al. J. Biol. Chem.* **273**, 11017–11024 (1998).
9. Gustavsson, J. *et al. FASEB J.* **13**, 1961–1971 (1999).
10. Levkowitz, G. *et al. Mol. Cell* **4**, 1029–1040 (1999).

Daedalus

Vitamins and minerals

Bottled mineral water, says Daedalus, is a pretentious adjunct to modern living. It is mainly water with a bit of fizz, but each bottle lovingly details its mineral analysis. Sometimes it lists traces of beneficial substances such as calcium or magnesium; often it records ones which are definitely harmful, such as sodium. But the implied magic of bottled water lies in its origin — some special spring or source far from urban contamination. One brand even claims to have been melted from glaciers laid down centuries ago. Oddly enough, the magic of mineral water (unlike its chemical composition) fades rapidly. Almost all bottles have a well-marked sell-by date. Consumers of this potent elixir gladly pay the vast price demanded for it. It lets them avoid alcohol in restaurants and pubs without appearing too poor or too mean to pay lavishly for a drink.

Daedalus now plans to lend a little scientific credibility to this profitable product. He notes the vast trade in mineral and vitamin supplements. Logically, these are quite a cheap form of health insurance. Vitamin deficiencies can sometimes occur; and the lack of elements such as selenium and zinc can cause medical problems in susceptible people. And coronary heart disease, at least in Britain, seems to occur more in regions with soft water. Hard water, containing calcium and magnesium, may have a significant protective effect.

So DREADCO's 'Reinforced Water' (slogan: *Harder than hard!*) will combine the chic, magic and high price of mineral water with the genuine benefits of diet supplements. Its main active ingredients will be calcium and magnesium, of which we need about 0.8 gram and 0.3 gram a day respectively, and vitamin C, of which 0.1 gram a day is not excessive but which is unpleasantly acid. To force these palatably into solution, Reinforced Water will be pressurized with several atmospheres of carbon dioxide. It will then take up calcium and magnesium copiously as bicarbonates, and vitamin C as neutral calcium ascorbate. The other vitamins and minerals, needed in mere milligrams or less, should dissolve easily.

Smart, vigorously fizzy, powerfully healthful, but free of the pharmaceutical overtones of diet-supplement pills, Reinforced Water should dominate the mineral-water market. But dare DREADCO abandon mineral magic for mere scientific rigour? The company's marketeers are thinking of claiming that it issues from a spring on Mount Olympus, home of the Immortals.

David Jones

Immunocontraception of African elephants

A humane method to control elephant populations without behavioural side effects.

Concerted efforts to nurture elephant populations have resulted in elephant overpopulation in several areas, which in turn has led to damaging levels of browsing. As an alternative to culling entire family groups in order to control this damage, we have developed an immunocontraceptive vaccine from pig zona pellucida which safely and successfully controls free-roaming African elephants.

Immunocontraceptive vaccines cause the immune system to produce antibodies that prevent fertilization, without the side effects of hormonal contraceptives. The vaccine antigens are the proteins of the zona pellucida, the clear protein coat surrounding mammalian eggs. The surface structures of the elephant zona pellucida are very similar to those of the pig zona pellucida (pZP)^{1–3}.

Female zoo elephants vaccinated with pZP and an adjuvant all developed antibodies that persisted for 12–14 months^{1,2}, at a level equivalent to those found in horses given immunocontraception^{4,5}. Based on this, we planned field trials in Kruger National Park, in conjunction with the South African National Parks.

Initial trials using 41 adult female elephants tested the efficacy of pZP as an immunocontraceptive. Elephants were located from a helicopter, and females to be anaesthetized (by aerial darting) were identified as non-pregnant by the presence of a calf smaller than 1 metre high. We then used ultrasound scans to confirm that females were not pregnant, and all non-pregnant animals were bled to obtain pre-vaccination serum samples. Twenty-one elephants were given an initial vaccination of pZP with adjuvant; 20 controls received a placebo.

All treated elephants were fitted with radiocollars. The control females were fitted with numbered collars and paired in a family unit with a vaccinated elephant. The vaccinated elephants were located 6 weeks later and received a first booster, followed by another 6 months later. Both boosters were administered remotely with drop-out darts from a helicopter.

Twelve months after the initial vaccination, the elephants were recaptured and scanned for pregnancy. Of those treated with pZP, 19 were recaptured (two were not found because their radiocollars failed). Nine of the 19 were pregnant, ten were not. One of the pregnant elephants was in the last trimester of gestation (22 months) and gave birth to a healthy calf, showing that the vaccination of a pregnant elephant with

pZP has no effect on gestation, the fetus or parturition. Eighteen of the 20 control elephants were located, 16 of which were pregnant. Therefore, significantly (χ^2 , $P=0.005$) fewer vaccinated elephants (44%, 8/18) were pregnant than the control females (89%, 16/18).

Subsequently, we vaccinated ten elephants using a revised schedule. Females received an initial vaccination, followed by identical boosters delivered from a helicopter two and four weeks later. All elephants were fitted with radiocollars; five had global positioning satellite (GPS) collars (Lotek, Newmarket, Canada) which recorded their location hourly.

Of the ten elephants, two (20%) were pregnant after 10 months. This was significantly (χ^2 , $P=0.001$) lower than the conception rate of the control elephants (89%, 16/18) and initial immunocontraception rates (44%, 8/18).

Female elephants with oestradiol implants have shown aberrant behaviour by separating off within the family unit (D.G., personal observation). GPS-collar location data indicated that there was no abnormal separation of the vaccinated females within a family unit over 8 months. This, combined with field observations of vaccinated females, suggests that the immunocontraceptive vaccine causes no behavioural abnormalities.

Finally, we tested the reversibility of immunocontraception, and its application for a second consecutive year. Of seven elephants from the group that had initially received immunocontraception, four were vaccinated with pZP and adjuvant, and

three were not. Twelve months later, the seven elephants were captured and re-evaluated. Ultrasound scans showed that all three untreated females had conceived again, compared with none of the vaccinated elephants, although all were cycling. This indicates that the vaccine is reversible, and that it has no deleterious effect on the ovary and its cyclicity.

Elephants are intelligent and empathetic mammals, and culling is a last resort in controlling their numbers. Our immunocontraceptive study shows that free-roaming African elephants vaccinated with pZP are protected against conception. This pZP immunocontraception is safe and reversible and is thus a practical tool for controlling elephant populations.

R. A. Fayrer-Hosken*, **D. Grobler†**,
J. J. Van Altena†, **H. J. Bertschinger‡**,
J. F. Kirkpatrick§

*College of Veterinary Medicine, University of Georgia, Athens, Georgia 30602-7385, USA
e-mail: rfh@calc.vet.uga.edu

†Kruger National Park, Private Bag X402, 1350 Skukuza, South Africa

‡Department of Theriogenology, University of Pretoria, Private Bag X04, 0110 Onderstepoort, South Africa

§ZooMontana, 2100 South Shiloh Road, Billings, Montana 59106, USA

1. Fayrer-Hosken, R. A. *et al. Theriogenology* **47**, 397 (1997).

2. Fayrer-Hosken, R. A. *et al. Theriogenology* **52**, 835–846 (1999).

3. Fayrer-Hosken, R. A., Bertschinger, H. J., Kirkpatrick, J. F., Turner, J. W. & Liu, I. K. M. *Bulletin* **25**, 18–21 (1997).

4. Willis, L. P., Heusner, G. L., Warren, R. J., Kessler, D. & Fayrer-Hosken, R. A. *J. Eq. Vet. Sci.* **14**, 364–370 (1994).

5. Kirkpatrick, J. F., Turner, J. W., Liu, I. K., Fayrer-Hosken, R. & Rutberg, A. T. *Reprod. Fertil. Dev.* **9**, 105–110 (1997).

Evolutionary biology

Sexual conflict and speciation

Sexual conflict occurs because males are selected to produce as many offspring as possible, even if this means lowering the overall reproductive output of individual females. A new model proposed by Gavrilits¹ suggests that strong asymmetries between males and females in the costs and benefits of mating will create runaway coevolution between the sexes, promoting rapid divergence between populations and hence speciation. This is an intriguing possibility, not least because it runs counter to existing models² which suggest that greater sexual conflict will result in males mating

more indiscriminately, breaking down reproductive barriers between divergent populations. One reason for this difference is that the new model is based on the idea that females can avoid costs of mating if they are incompatible with some males, whereas we suggest that in reality this may rarely be the case.

Gavrilits' model¹ assumes a quadratic relation between female fitness and the proportion of the male population with which she is compatible (morphologically, physiologically or genetically). The shape of this relationship is not theoretically derived (as it might be), but is the simplest function under which there is an intermediate optimum proportion of compatible males. This relationship is based on observations suggesting that females experiencing unusually

high or low mating rates may have reduced fitness³, but such analyses must be treated with caution. Females are expected to be adapted to the number of matings they experience⁴, even if this is a compromise between their optimal mating rate and that which males attempt to impose.

More significantly, most of the costs to females that Gavrillets sees as driving sexual conflict (such as predation, sensory exploitation, different costs of mating or seminal fluid toxicity)¹ are paid as a result of matings *per se*, and hence can only be prevented if matings themselves are avoided. However, if incompatibilities prevent mating, then males are expected to be able to determine rapidly whether a particular female is compatible. Hence incompatibilities that prevent mating between particular male and female phenotypes will simply result in males only attempting to mate with females with whom they are compatible, with no reduction in the number of matings or attempted matings to which each female is subjected.

This suggests that the potential for sexual conflict to promote divergence may be limited to species without the possibility of pre-copulatory mate choice, but in which there are still costs to being compatible with too many males. There may be examples of such species — for instance, broadcast spawners may suffer from polyspermy — but such cases are likely to be rare.

Although we question the generality of Gavrillets' model¹, it does indicate the potential for the different priorities of males and females to drive evolution of reproductive isolation. Perhaps rather than lumping so much biology together under the title 'sexual conflict', we need to consider the specific processes involved.

Tom Tregenza, Roger K. Butlin, Nina Wedell
School of Biology, University of Leeds,
Leeds LS2 9JT, UK
e-mail: gentbt@leeds.ac.uk

1. Gavrillets, S. *Nature* **403**, 886–889 (2000).
2. Parker, G. A. & Partridge, L. *Phil. Trans. R. Soc. Lond. B* **353**, 261–274 (1998).
3. Arnqvist, G. & Nilsson, T. *Anim. Behav.* (in the press).
4. Holland, B. & Rice, W. R. *Proc. Natl Acad. Sci. USA* **96**, 5083–5088 (1999).

Gavrillets replies — Tregenza *et al.* maintain that my model's predictions¹ run counter to the model of Parker and Partridge², but it is not straightforward to compare these two classes of model because of their inherent differences.

First, my model considers genetic divergence of allopatric populations, whereas the models developed in ref. 2 describe the evolutionary consequences of a secondary contact between populations that have already diverged: that is, they model reinforcement of reproductive isolation rather than its emergence. Second, Parker and Partridge² use an evolutionarily stable

strategy framework with a set of fixed evolutionary options, whereas I model continuous coevolution of the sexes, during which evolutionary options are continuously changing.

In spite of these differences and contrary to the claim of Tregenza *et al.*, there is no contradiction in the predictions of both types of model. Parker and Partridge propose that mating conflict could be either a hindrance to isolation if 'male-win' scenarios prevail, or a facilitator if females tend to win — exactly as predicted by my model if coevolution is restricted. However, my model takes an additional step by considering the possibility of continuous coevolution of the sexes. In this case the prediction is that neither sex will win the sexual conflict, but rather that there will be a dynamic coevolutionary compromise.

Tregenza *et al.* question the generality of my model's assumption that females have an intermediate optimum mating rate, as well as the specific (quadratic) function I used to model the relation between a female's fitness and her mating rate: in fact, the model is well supported by insect³ and other data^{4,5}, and the idea that excessive mating rates are bad for females is the essence of sexual conflict.

Inadequate mating rates are also detrimental⁶. Thus, with sexual conflict the optimum mating rate must be intermediate. As for the shape of the relation between female mating costs and mating rate, more data are indeed necessary. However, I do not anticipate that using functions more complicated than a quadratic one will affect my main conclusions¹.

Tregenza *et al.* suggest that, in most species, males will rapidly identify and attempt to mate only with those females with whom they are compatible, thus reducing the potential for sexual conflict to promote divergence. But even if the necessary 'indicators' of female compatibility should be readily available and the males smart enough to exploit them, in a polymorphic population different females will experience a variable number of matings. This will induce fitness differences and initiate runaway coevolution, resulting in genetic divergence of isolated populations. Although the model does not describe all the specific processes involved in sexual conflict, other weaknesses are not yet apparent.

S. Gavrillets

Department of Ecology and Evolutionary Biology,
Department of Mathematics, University of
Tennessee Knoxville, Tennessee 37996, USA

1. Gavrillets, S. *Nature* **403**, 886–889 (2000).
2. Parker, G. A. & Partridge, L. *Phil. Trans. R. Soc. Lond. B* **353**, 261–274 (1998).
3. Arnqvist, G. & Nilsson, T. *Anim. Behav.* (in the press).
4. Fowler, K. & Partridge, L. *Nature* **338**, 760–761 (1989).
5. Gens, D. & Riddle, D. L. *Nature* **379**, 723–725 (1996).
6. Jennions, M. D. & Petrie, M. *Biol. Rev.* **75**, 21–64 (2000).

Sandwich films

Properties of a new soft magnetic material

The development of advanced electromagnetic devices has been constrained by a lack of soft magnetic materials with a suitably high saturation magnetization (over 20 kilogauss) and a large permeability roll-off frequency (greater than 1 gigahertz). For example, magnetic hard-disk-drive technology is rapidly approaching the perceived superparamagnetic limit at which the stored bits become thermally unstable¹ — disks with higher anisotropy are more stable but are not usable because magnetic write heads become saturated. Here we describe a new soft magnetic material with a saturation magnetization of 24 kilogauss and a large permeability of 1,000–1,400 in a wide frequency range of up to about 1.2 gigahertz. This new material promises to have wide application in devices such as magnetic recording heads and integrated inductors.

Soft magnetic materials such as $\text{Co}_{0.57}\text{Ni}_{0.13}\text{Fe}_{0.30}$ (ref. 2) and FeAlN (ref. 3) have a saturation magnetization of about 20 kG, and $\text{Fe}_{1-x}\text{Co}_x$ alloys (where $0.3 \leq x \leq 0.4$) have the highest saturation magnetization that occurs in nature (about 24.5 kG). However, the coercivity of these FeCo alloys tends to be greater than 5 oersteds, making them unsuitable for magnetic write heads⁴. Attempts to obtain soft magnetic materials have involved making artificial structures⁵.

We have created a sandwich structure, $\text{Ni}_{0.81}\text{Fe}_{0.19}/(\text{Fe}_{0.7}\text{Co}_{0.3})_{0.95}\text{N}_{0.05}/\text{Ni}_{0.81}\text{Fe}_{0.19}$, in which the FeCoN film is 100 nm thick, and each $\text{Ni}_{0.81}\text{Fe}_{0.19}$ Permalloy layer is 5 nm thick, comprising only 4.5% of the volume of the sandwich. Typical hysteresis loops (magnetization, M , plotted against magnetic field, H) of the sandwich films, determined by using a vibrating sample magnetometer, are shown in Fig. 1, inset:

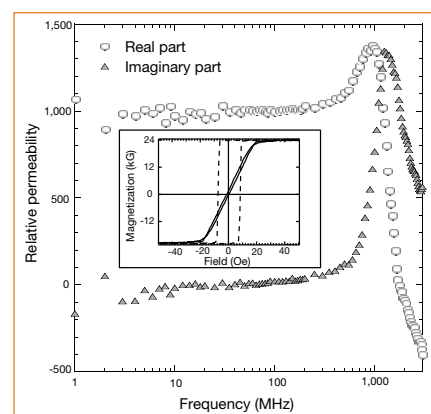


Figure 1 Relative permeability, both real and imaginary parts, versus frequency for a typical $\text{Ni}_{0.81}\text{Fe}_{0.19}/(\text{Fe}_{0.7}\text{Co}_{0.3})_{0.95}\text{N}_{0.05}/\text{Ni}_{0.81}\text{Fe}_{0.19}$ (5 nm)/(100 nm)/(5 nm) film. Inset, typical $M-H$ loops of the sandwich film. Dashed line, easy axis loop; solid line, hard axis loop.

coercivities of the sandwich films are 0.6 and 7.8 Oe in the hard and easy axis, respectively; the anisotropy field, H_k , is about 20 Oe. Much thicker soft magnetic films with similar properties can be realized by laminating many $\text{Ni}_{0.81}\text{Fe}_{0.19}/(\text{Fe}_{0.7}\text{Co}_{0.3})_{0.95}\text{N}_{0.05}$ bilayers.

In contrast, coercivities of the FeCoN single layer are about 5 and 18 Oe in the hard and easy axis, respectively. A purely magnetostatic consideration would give the average coercivity of the sandwich as about 91% of that of the FeCoN single layer if the coercivities of Permalloy are negligible. Therefore, the observed reduction in coercivity due to sandwiching is surprisingly large. We believe that the reduction of magnetic anisotropy dispersion due to the Permalloy seed and applied deposition field (about 50 Oe) may be important here.

We also measured the hard-axis magnetic permeability, both real and imaginary parts, of the new material using a permeance meter with a frequency range up to 3 GHz (ref. 6) (Fig. 1). The real permeability is 1,000–1,400 at frequencies up to ~1.2 GHz, and the imaginary permeability peaks at ~1.5 GHz, which corresponds to the ferromagnetic resonance frequency, and is slightly lower than the calculated one: $f_{\text{FMR}} = \gamma \sqrt{4\pi M_s \times H_k} = 1.9$ GHz, where $\gamma = 2.8$ MHz per oersted and is the gyromagnetic ratio, and $4\pi M_s$ is the saturation magnetization, all in gaussian units. The resistivity of the film is ~50 $\mu\Omega$ cm, and its cut-off frequency due to eddy currents is ~8.6 GHz, so the eddy currents can be neglected.

The permeability spectra confirm that our new films are promising for use in extremely high-density magnetic write heads³ as well as in integrated inductors operating in the gigahertz range⁷. Other soft magnetic materials described for applications at gigahertz frequencies, such as $\text{Co}_{0.443}\text{Fe}_{0.191}\text{Hf}_{0.145}\text{O}_{0.221}$ (ref. 8) and $\text{Fe}_{0.878}\text{Cr}_{0.046}\text{Ta}_{0.002}\text{N}_{0.074}$ (ref. 9), have a permeability of about 200 and a lower saturation magnetization.

S. X. Wang*, N. X. Sun*, M. Yamaguchi†, S. Yabukami†

*Laboratory of Advanced Materials, Department of Materials Science and Engineering, Stanford University, Stanford, California 94305-4045, USA

†Research Institute of Electrical Communication, Tohoku University, Sendai 980-8577, Japan
e-mail: sxwang@ee.stanford.edu

1. Charap, S. H., Lu, P. L. & He, Y. *IEEE Trans. Magn.* **33**, 978–983 (1997).

2. Osaka, T. et al. *Nature* **392**, 796–798 (1998).

3. Wang, S. X. & Taratorin, A. M. *Magnetic Information Storage Technology* (Academic, Boston, 1999).

4. Sun, N. X. & Wang, S. X. *IEEE Trans. Magn.* (in the press).

5. Clow, H. *Nature* **194**, 1035–1036 (1962).

6. Yabukami, S. et al. *J. Appl. Phys.* **85**, 5148–5150 (1999).

7. Yamaguchi, M. et al. *J. Appl. Phys.* **85**, 7919–7922 (1999).

8. Hayakawa, Y. et al. *J. Appl. Phys.* **81**, 3747–3752 (1997).

9. Jin, S. et al. *J. Appl. Phys.* **70**, 3161–3163 (1997).

Virology

KSHV-like herpesviruses in chimps and gorillas

Among the herpesviruses¹, KSHV (Kaposi's-sarcoma-associated herpesvirus) is the human prototype of the rhadinovirus genus². Rhadinoviruses (or γ_2 -herpesviruses) are found in several animal species, including New and Old World monkeys, but not in the great apes^{3–5}. Here we describe the detection and sequencing of a polymerase gene fragment from three new rhadinoviruses discovered in chimpanzees and in a gorilla, which are more closely related to KSHV than to any other virus of this genus described so far. Our results indicate that the great apes from central Africa

could provide a reservoir of new γ_2 -herpesviruses that are potentially transmissible to humans.

As KSHV and Kaposi's sarcoma are endemic in central Africa, we tested for KSHV-related viruses in great apes from central Africa. Blood specimens were taken from 30 wild-born animals (25 chimpanzees, *Pan troglodytes*, and five gorillas, *Gorilla gorilla*) originating from Cameroon and Gabon⁶. An immunofluorescence assay that detects both latent and lytic KSHV antigens revealed a clear reactivity in the plasma of 20 out of 25 chimpanzees and two out of five gorillas at a 1 in 40 dilution⁷.

We amplified a fragment of DNA isolated from peripheral blood mononuclear cells taken from five of these animals (four chimpanzees and one gorilla) by using nest-

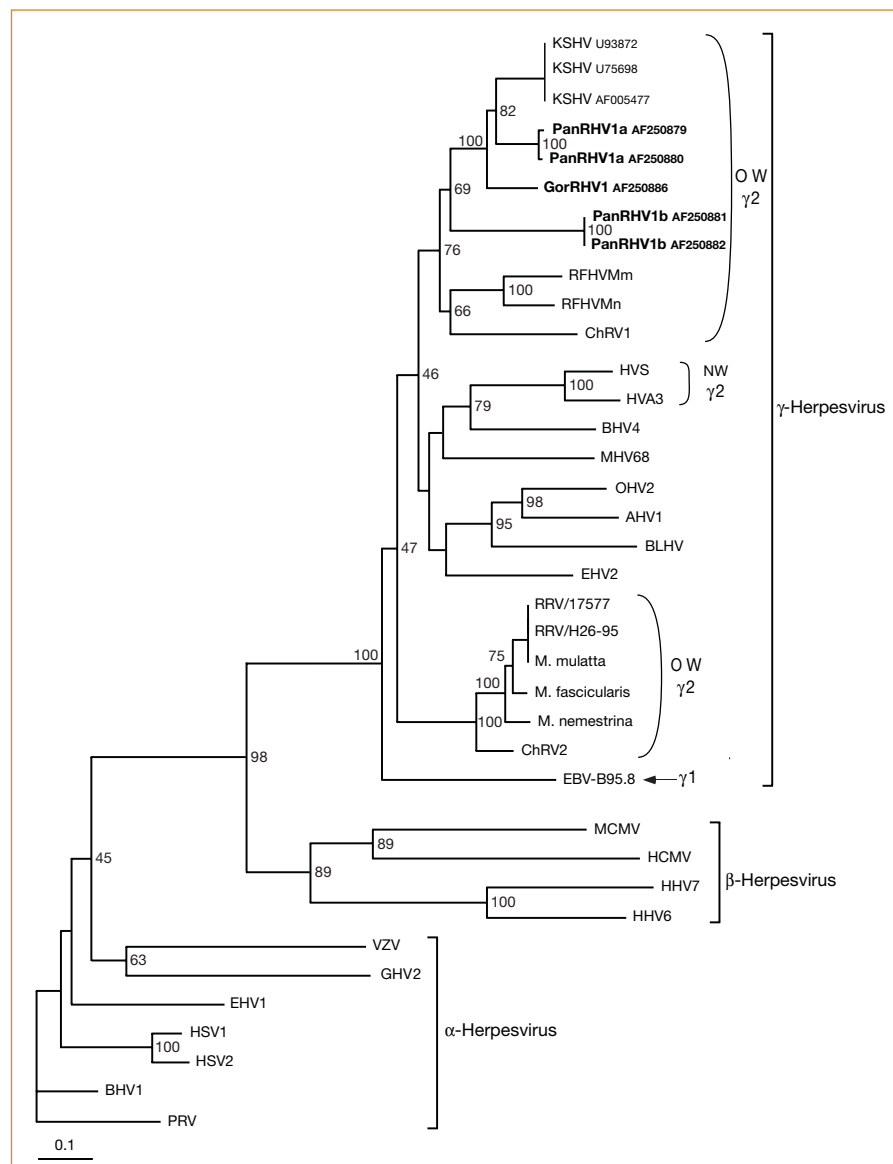


Figure 1 Phylogenetic tree resulting from analysis of 37 selected 454-base-pair fragments⁵ of the herpesvirus DNA-polymerase gene. Phylogeny was derived by using the neighbour-joining method applied to pairwise sequence distances calculated by the Kimura two-parameter method. Horizontal branch lengths are drawn to scale, with the bar indicating 0.1 nucleotide replacements per site. Numbers at each node indicate the percentage of bootstrap samples (out of 100) in which the cluster to the right is supported. Brackets on the right indicate previously defined subfamily and genus herpesviral classification. OW, Old World; NW, New World.

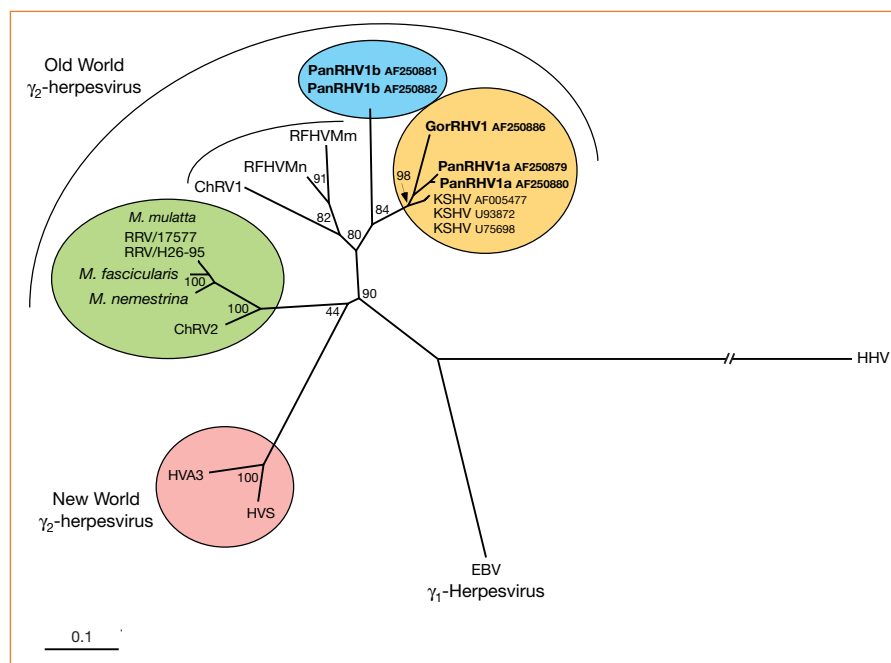


Figure 2 Neighbour-joining protein distance tree for the 151 amino-acid residues encoded by a 454-base-pair fragment⁵ of the DNA-polymerase gene. Sequences were aligned by using ClustalW and analysed with PROTDIST and NEIGHBOR programs in PHYLIP. One hundred replica samplings were subjected to bootstrap analysis (SEBOOT). Branch lengths are proportional to the evolutionary distance (scale bar) between the taxa. GenBank accession numbers for the new chimp and gorilla KSHV-like sequences are respectively AF250879, AF250880, AF250881, AF250882 and AF250886.

ed polymerase chain reaction and degenerate primers⁵; this was 476 base pairs long and encoded a fragment of the conserved herpesvirus DNA polymerase. Comparative analysis with all the other available related herpesvirus sequences (Figs 1, 2) revealed the existence of two new chimpanzee KSHV-like viruses (which we named PanRHV1a and PanRHV1b, for pan-rhadinoviruses 1a and 1b) and a new gorilla KSHV-like virus (GorRHV1, for gorilla rhadinovirus 1). These new sequences from great apes appear to be outside the natural range of KSHV-sequence variations found in humans.

Phylogenetic analysis by different methods placed these three viruses in the rhadinovirus genus, which, with the KSHV, the macaque RFHVMm and RFHVMn, and African green monkey ChRV1 strains, forms a new clade of Old World viruses that is supported by high bootstrap values (Fig. 1).

A second analysis, restricted to all the available primate rhadinovirus DNA-polymerase genes (Fig. 2), indicated that there are five separate lineages among these γ_2 -herpesviruses, one for the New World monkeys (HVS and HVA3), two for the macaques (comprising RFHVMm, RFHVMn, RRV, *Macaca mulatta*, *M. nemestrina* and *M. fascicularis*) and the African green monkeys (ChRV2 and ChRV1), one for the new PanRHV1b strains, and one that includes KSHV and the new PanRHV1a and GorRHV1 strains. Among the chimpanzees, none of the new strains seems to be associated with a genetic

subspecies as they were found to occur nonspecifically in *Pan troglodytes vellerosus* and *Pan troglodytes troglodytes* by mitochondrial-DNA analysis.

The increased contact between great apes and humans in central Africa generated by hunting and butchering for local meat and/or the 'bushmeat trade', as well as by recent socio-economic change such as intense commercial logging, provides more opportunity for interspecies viral transmission⁸. Such infectious events, although rare, may lead to a dead-end infection, as reported for foamy viruses after monkey bites⁹. Alternatively, and more probably in the case of herpesviruses, they may cause persistent infection, with a possible founder effect in restricted populations, leading under particular circumstances to endemic or epidemic dissemination within larger human populations.

The two human retroviruses HIV-1 and HTLV-1, at least some subtypes of which probably originated from the central African chimpanzee viruses SIV/CPZ and STLV-1, provide two examples of infection that has been zoonotically transmitted in this way, albeit with a different epidemiological fate and pathogenicity in humans^{6,10,11}. Our discovery of the new KSHV-related rhadinovirus sequences in chimpanzees and gorillas will enable their likely human counterparts to be identified.

Vincent Lacoste*, Philippe Maucière*,
Guy Dubreuil†, John Lewis‡,
Marie-Claude Georges-Courbot‡,
Antoine Gessain*

*Unité d'Oncologie Virale, Département du SIDA et des Rétrovirus, Institut Pasteur, 25–28 rue du Dr Roux, 75724 Paris Cedex 15, France

e-mail: agessain@pasteur.fr

†Centre Pasteur du Cameroun, BP 1274, Yaoundé, Cameroon

‡Centre International de Recherches Médicales, Franceville, Gabon

§International Zoo Veterinary group, Keighley, West Yorkshire BD21 1AG, UK

1. Roizmann, B. et al. *Arch. Virol.* **123**, 425–449 (1992).
2. Chang, Y. et al. *Science* **266**, 1865–1869 (1994).
3. Greensill, J. et al. *J. Virol.* **74**, 1572–1577 (2000).
4. Desrosiers, R. C. et al. *J. Virol.* **71**, 9764–9769 (1997).
5. Rose, T. M. et al. *J. Virol.* **71**, 4138–4144 (1997).
6. Corbet, S. et al. *J. Virol.* **74**, 529–534 (2000).
7. Chatlynne, L. G. et al. *Blood* **92**, 53–58 (1998).
8. Hahn, B. H. et al. *Science* **287**, 607–614 (2000).
9. Heneine, W. et al. *Nature Med.* **4**, 403–407 (1998).
10. Gao, F. et al. *Nature* **397**, 436–441 (1999).
11. Mahieux, R. et al. *J. Virol.* **72**, 10316–10322 (1998).

Addendum

Fish do not avoid survey vessels

P. G. Fernandes, A. S. Brierley, E. J. Simmonds, N. W. Millard, S. D. McPhail, F. Armstrong, P. Stevenson, M. Squires

Nature **404**, 35–36 (2000)

Our title has led some readers to misinterpret our conclusions. Questions have also been raised regarding the speed at which the observations were carried out.

Autosub-1, the Autonomous Underwater Vehicle that we compared with the fisheries research vessel *Scotia*, had an operating speed of 3 knots. Vessel noise is usually thought to increase with speed, but the noise signature of *Scotia* (Fig. 1) shows that she makes a similar amount of noise at a low speed and at the normal survey speed of 10 knots. Our conclusions are, therefore, valid at the usual survey speed for this vessel.

Our paper's title has, in some cases, been taken to imply that avoidance is not a problem for any research vessel, without considering the qualifications we thought had been made explicit in the text. The *Scotia* is very quiet: she was the first vessel to be built to the International Council for the Exploration of the Sea's specifications for limiting noise¹ (Fig. 1). We believe that our findings, valid only for such noise-reduced vessels, endorse this objective and justify the 5–10% increase in build cost. Our research shows that fisheries research vessels should be built to the same noise specification as *Scotia* in order to ensure that avoidance is not a source of bias.

1. Mitson, R. B. *ICES Co-op. Res. Rep.* **209**, 1–61 (1995).

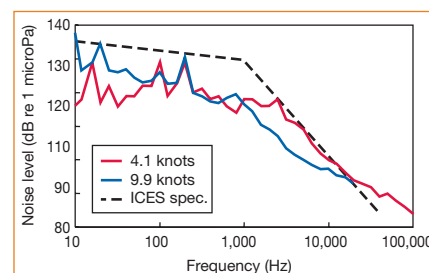


Figure 1 Noise levels of the *Scotia* at different speeds compared with the noise specification recommended by the International Council for the Exploration of the Sea.

Compartmental specificity of cellular membrane fusion encoded in SNARE proteins

James A. McNew⁺, Francesco Parlati[†], Ryouichi Fukuda^{*,†}, Robert J. Johnston, Keren Paz, Fabienne Paumet, Thomas H. Söllner & James E. Rothman

Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, Box 251, New York, New York 10021, USA

[†] These authors contributed equally to this work

Membrane-enveloped vesicles travel among the compartments of the cytoplasm of eukaryotic cells, delivering their specific cargo to programmed locations by membrane fusion. The pairing of vesicle v-SNAREs (soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors) with target membrane t-SNAREs has a central role in intracellular membrane fusion. We have tested all of the potential v-SNAREs encoded in the yeast genome for their capacity to trigger fusion by partnering with t-SNAREs that mark the Golgi, the vacuole and the plasma membrane. Here we find that, to a marked degree, the pattern of membrane flow in the cell is encoded and recapitulated by its isolated SNARE proteins, as predicted by the SNARE hypothesis.

Accuracy in membrane fusion is vital for the growth and division of cells and for propagating the spatial organization of biochemical events within them. But how compartmental specificity is achieved remains unknown. The simplest possibility is that it is encoded in the fusion mechanism itself¹. The pairing of v-SNAREs and t-SNAREs^{2–4} between bilayers is the active principle of intracellular membrane fusion^{2,5–9}. The cytoplasmic domains of these integral membrane proteins self-assemble into a rod-like SNARE complex^{10,11} that docks vesicles to target membranes. The assembly of such ‘SNAREpins’ (also termed *trans*-SNARE complexes) between artificial lipid bilayers triggers spontaneous fusion².

Fusion mediated by isolated SNAREs occurs efficiently^{2,5} and is driven energetically by the exceptional stability of the SNARE complex^{11,12}, which is used to do work on the lipid bilayers¹³. SNARE-dependent fusion involves both monolayers equally² and the mixing of the aqueous contents⁶. When an auto-inhibitory domain¹⁴ is removed, fusion by isolated SNAREs⁵ can proceed as fast as overall exocytosis in cells⁷. Isolated SNAREs drawn from several pathways, including transport from endoplasmic reticulum (ER) to the Golgi⁸, homotypic fusion of vacuoles⁹ and regulated exocytosis², fuse lipid bilayers.

Altogether this body of work establishes that SNAREpins form the basis of intracellular membrane fusion and that fusion automatically follows after *trans*-SNARE pairing unless a special mechanism exists to prevent it, as in regulated exocytosis. Here, complete assembly of *trans*-SNARE complexes is blocked^{15,16}, explaining the extreme rapidity of the initial burst of exocytosis in synapses and certain other systems when exocytosis is triggered.

In cells, a given set of SNARE proteins can participate in fusion only if the following conditions are met. First, they must be localized in two different compartments. Second, the two compartments must be able to access each other. Third, particular combinations of heavy and light chains must localize to the same compartment to allow the assembly of a functional t-SNARE⁹. Fourth, the cytoplasmic domains of the SNAREs must be able to assemble to link the two membranes within the time frame of transport. Fifth, the membrane anchors¹³ and the linkers¹⁷ connecting to the cytoplasmic domains must be functional for fusion. Last, their anchoring

topology must be compatible with fusion⁸. In principle, compartmental specificity of membrane fusion could be encoded in SNARE proteins at any or all of these levels, of which only the final three involve properties that are retained in isolated proteins.

Here we provide the most direct test possible of whether, and to what extent, the compartmental specificity of intracellular membrane fusion is an inherent property encoded in the physical chemistry of the SNARE proteins themselves. For this purpose we have used SNARE proteins drawn from one cell, *Saccharomyces cerevisiae*, because all of its SNARE proteins have been identified from its completed genome sequence^{4,18}. We take advantage of the functional identification of the v- and t-SNAREs mediating the fusion of ER-derived transport vesicles with the Golgi⁸, the homotypic fusion of vacuoles⁹ and fusion with the plasma membrane^{2,19,20}. We tested these three t-SNAREs for fusion activity in all possible combinations with all the potential v-SNAREs of a yeast cell to establish the pattern of specificity encoded in the isolated SNARE proteins.

Fusion of vesicles by exocytic v- and t-SNAREs

The heterodimer of the syntaxin Sso1 (ref. 20) and the SNAP25 homologue Sec9 (ref. 19) is a t-SNARE of the yeast plasma membrane, and Snc2 (refs 21, 22; interchangeable genetically with Snc1) is its cognate v-SNARE. Sso1 shares 42% sequence identity with rat Syntaxin1A, whereas Snc2 has 50% identity with neuronal VAMP2. The yeast plasma membrane t-SNARE complex thus closely resembles its neuronal counterpart².

We expressed full-length versions of Snc2 and Sso1 as His-tagged proteins in *Escherichia coli*. We also expressed the region of Sec9 homologous to SNAP25 (Sec9c)²³ as a glutathione S-transferase (GST) fusion protein. The t-SNARE complex was reconstituted by mixing His₈-Sso1 with GST-Sec9c in octylglucoside followed by lipid addition, dilution and dialysis. After isolation and recovery of the liposomes by flotation, the GST moiety was removed from Sec9c by cleavage with thrombin (Fig. 1a). About 60–70% of the GST-Sec9c was reconstituted so that it faced the outside of the liposome and could be cleaved by thrombin; the remainder was resistant. However, the GST does not affect the activity of Sec9c (data not shown). Fusion was assayed by lipid mixing as described².

Fluorescent ‘donor’ liposomes bearing the v-SNARE Snc2 spontaneously and efficiently fuse with the non-fluorescent ‘acceptor’

⁺ Present addresses: Department of Biochemistry and Cell Biology, Rice University, MS140, 6100 Main Street, PO Box 1892, Houston, Texas 77251-1892, USA (J.A.M.); Department of Biotechnology, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku Tokyo 113-8657, Japan (R.F.).

liposomes bearing the t-SNARE Sso1/Sec9c (Fig. 2); and progressively lowering the number of copies of the v-SNARE Snc2 per liposome lowers the rate of fusion. The free cytoplasmic portion of the v-SNARE Snc2 inhibits fusion, confirming a requirement for SNARE complex assembly. The rate of fusion of the cognate yeast SNAREs is about three times faster than cognate neuronal SNAREs measured in parallel (data not shown). Liposomes containing the exocytic/neuronal v-SNARE VAMP2 also readily fuse with t-Sso1/Sec9c liposomes, and liposomes bearing the exocytic/neuronal t-SNARE syntaxin1A/SNAP25 fuse with those bearing Snc2. Although these fusion reactions occur at the slower rate of the cognate exocytic/neuronal SNAREs (data not shown), they nonetheless indicate that SNARE pairing of the plasma membrane is highly conserved in evolution.

Fusion specificity encoded in v- and t-SNAREs

Altogether, three cognate pairs of v- and t-SNAREs have now been

identified according to their function in bilayer fusion: v-Bet1 with t-Sed5/Bos1, Sec22 (ER–Golgi)⁸; v-Nyv1 with t-Vam3/Vam7, Vti1 (vacuole fusion)⁹; and v-Snc2 with t-Sso1/Sec9c (Golgi–plasma membrane; shown here). To test the compartmental specificity of SNARE-dependent fusion, we tested each of these three v-SNAREs with each of the three t-SNAREs for a total of nine possible combinations (Fig. 3). The three known cognate matches are shown in Fig. 3a, e and i (shaded panels).

Of the six other mixtures, only one fused significantly (Fig. 3b, lightly shaded). This was a match between the vacuolar v-SNARE and the plasma membrane t-SNARE, representing the ubiquitous pathway of lysosomal secretion, not yet shown to occur in yeast. This fusogenic combination therefore seems likely to represent another physiologically relevant, cognate match.

The vesicles used in Fig. 3 had concentrations of v- or t-SNARE proteins (see Fig. 1b and Table 1) that were as similar as we could achieve. The rate of fusion in Fig. 3a (v-Snc2 with t-Sso1/Sec9c) corresponds to the lowest rate in Fig. 2 (filled diamonds) because the lowest (0.25×) concentration of Snc2 more closely matched the concentration of the other v-SNAREs tested in Fig. 3. It should be noted that Snc2 stains poorly compared with other SNARE proteins (Fig. 1).

The genome of *S. cerevisiae* encodes a total of 14 non-syntaxin SNAREs¹⁸ that could potentially function as either v-SNAREs or light chains of t-SNAREs, or both (that is, as the cognate v-SNARE for one t-SNARE and as the light chain of another).

Four of the non-syntaxin SNAREs do in fact function as cognate v-SNAREs, namely Snc1 and Snc2 (Golgi–plasma membrane), Bet1 (ER–Golgi transport) and Nyv1 (vacuolar fusion), on the basis of their ability to promote fusion by binding a t-SNARE located in a different bilayer. As Snc1 and Snc2 are largely interchangeable genetically²⁴ and are 77% identical, we expected that Snc1 (like Snc2) might function as a v-SNARE with the plasma membrane t-SNARE Sso1/Sec9c; and this is confirmed in Fig. 4c.

The remaining 10 non-syntaxin SNAREs are Ykt6, Sec22, Bos1, Gos1, Sft1, Tlg1, Vti1, Vam7, Sec9 and Spo20. All have been shown

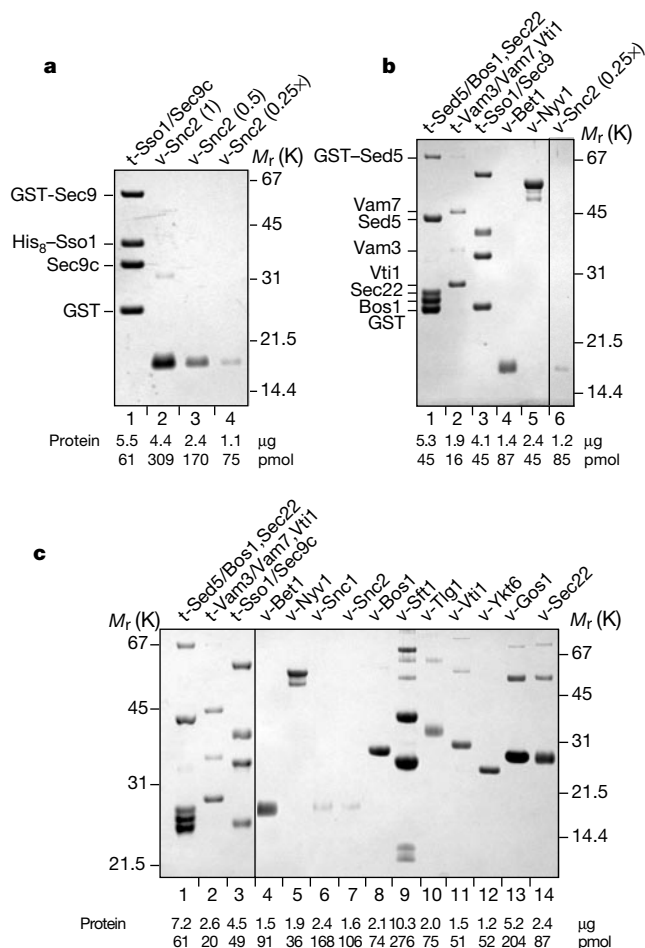


Figure 1 Characterization of liposomes. **a**, Coomassie-stained gel showing the liposomes used in Fig. 2. The GST–Sec9 that remains after thrombin cleavage is the result of incorrectly orientated (lumen-facing) t-SNARE complexes. This protein is completely cleaved when the membrane is dissolved by the addition of detergent during digestion (data not shown). Free GST remains in the liposome preparation after thrombin digestion. **b**, Coomassie-stained gel showing the liposomes used in Fig. 3. **c**, Coomassie-stained gel showing the liposomes used in Fig. 4. Five microlitres of each liposome preparation were loaded, except in the case of Snc1 and Snc2 where 10 μ l were loaded owing to poor staining with Coomassie blue (**b**, lanes 4 versus 5). The minor bands greater than 45K in lanes 9–14 are probably multimers that are resistant to SDS denaturation. The protein pattern in lane 9 (Sft1) is derived from GST–Sft1 cleavage in liposomes. The ~35K band represents GST–Sft1 that is facing the vesicle lumen, the 25K band is GST, and the ~12K band is Sft1.

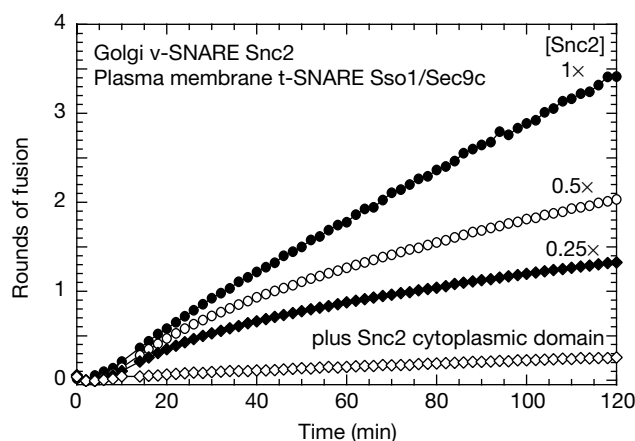


Figure 2 The yeast exocytic SNAREs can fuse vesicles. **a**, Sso1, Snc2 and GST–Sec9 were expressed and purified from *E. coli*. Sso1 (~212 μ g) and GST–Sec9 (~2,075 μ g) were mixed (~1:5 molar ratio) and incubated overnight at 4 °C. The t-SNARE complex was reconstituted into unlabelled (acceptor) phospholipid vesicles. Snc2 (~270 μ g) was reconstituted into fluorescently labelled (donor) phospholipid vesicles (1×). Additional amounts (0.5×, ~135 μ g) and (0.25×, ~68 μ g) of Snc2 were reconstituted into donor liposomes. Donor v- and acceptor t-SNARE liposomes were mixed 1:9 (by volume) in a microtitre plate on ice. For soluble domain inhibition, t-SNARE liposomes were pre-incubated on ice for 15 min with competitor soluble domain GST–Snc2 Δ TMD (~36 μ g, ~990 nmol) where indicated. The plate was transferred to a 37 °C fluorescent plate reader and NBD fluorescence (excitation 460 nm, emission 538 nm) was monitored at 2-min intervals. At 120 min, detergent was added to determine NBD fluorescence at infinite dilution. Data are expressed as round of fusion measured as fold lipid dilution^{5,17}.

by biochemical and/or genetic criteria to form complexes with other SNAREs and to function *in vivo*⁴.

Because three of the non-syntaxin SNAREs (Vam7, Sec9 and Spo20) are not intrinsic membrane proteins and are not lipid anchored, they cannot function by themselves as v-SNAREs. However, this does not prevent them from serving as light chains of t-SNAREs. Vam7 is a light chain of the vacuolar t-SNARE⁹, and Sec9 is the light chain of the plasma membrane t-SNAREs (unlike mammalian SNAP25, Sec9 is not palmitoylated). Spo20 is a sporulation-specific homologue of Sec9 (ref. 25).

Of the remaining seven non-syntaxin SNAREs, six are classic membrane proteins, and one, Ykt6, is anchored to bilayers through an isoprenoid lipid covalently attached at its carboxy terminus, which is probably a C₂₀ geranylgeranyl group attached at a CAAX box²⁶. Prenylation (and therefore membrane attachment) is required for Ykt6 function which includes transport at the ER/Golgi interface (most probably retrograde transport)²⁶.

Sec22 and Bos1 are located in the ER and Golgi, where they function in anterograde and probably retrograde transport as well as intra-Golgi transport^{4,27,28}. Gos1 and Sft1 are probably restricted to the Golgi and function within this organelle^{29,30}. All four of these SNAREs can participate *in vivo* in a series of at least two distinct SNARE complexes^{4,28,31} containing the same syntaxin, Sed5; one of these complexes (the tetramer Sed5/Bos1/Sec22/Bet1) has been fully characterized and shown to mediate bilayer fusion⁸.

Tlg1 was originally classified as a syntaxin, but further analysis¹⁸ showed that it is more closely related to non-syntaxin SNAREs. This SNARE protein is located in the late Golgi and in endosomes, and functions at the interfaces of the plasma membrane, endosomes and the Golgi⁴. Vti1 participates *in vivo* in distinct SNARE complexes containing any one of several syntaxins: Sed5 (Golgi)^{32,33}; Tlg2 (late Golgi/endosomes)³⁴; Vam3 (vacuoles)³⁵; or Pep12 (pre-vacuoles)³⁴. Correspondingly, select Vti1 mutations can affect either Golgi or

endocytic compartments³³. Vti1 can act as a light chain of a vacuolar t-SNARE⁹; whether it functions as a light chain or as a v-SNARE elsewhere is currently unknown.

In summary, the yeast genome encodes 11 actual or potential v-SNAREs. We expressed, purified and reconstituted each of them into fluorescent donor vesicles (Fig. 1c). In the case of Ykt6, the expressed soluble protein was covalently linked through the cysteine of its CAAX box to a geranylgeranyl (C₂₀) isoprenoid-containing moiety¹³ to provide a close equivalent to its probable natural membrane anchor²⁶. We determined the surface densities of these potential v-SNAREs in the reconstituted vesicles (Table 2); notably, the concentration in vesicles of the cognate v-SNARE (indicated by the prefix 'v') was similar or less than that of the other candidate v-SNAREs, favouring detection of possible fusion in non-cognate mixtures.

We then tested the 11 donor vesicles in all of the 33 possible combinations for their capacity to fuse with non-fluorescent acceptor vesicles bearing 1 of the 3 established species of t-SNAREs, which mark the Golgi (Sed5/Sec22/Bos1; Fig. 4a), the vacuole (Vam3/Vti1; Fig. 4b) and the plasma membrane (Sso1/Sec9c; Fig. 4c) *in vivo*. The extent of fusion after 2 h was normalized in each case to

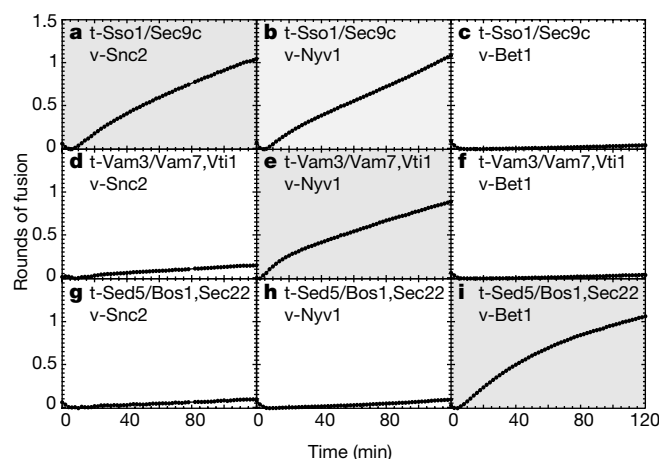


Figure 3 Compartmental specificity is determined by SNARE interactions. Liposomes containing the plasma membrane t-SNARE complex (t-Sso1/Sec9), the vacuolar t-SNARE complex (t-Vam3/Vam7,Vti1) and the ER–Golgi t-SNARE complex (t-Sed5/Bos1,Sec22) were independently reconstituted into liposomes. These t-SNARE liposomes were fused in all possible combinations with the v-SNAREs Snc2 (0.25x), Nyv1 and Bet1. Donor v- and acceptor t-SNARE liposomes were mixed 1:9 (5:45 μ l) in a microtitre plate on ice. The addition of 45 μ l of the t-SNAREs gave the following amounts of protein and lipid: t-Sso1/Sec9, 405 pmol protein, 102 nmol lipid; t-Vam3/Vam7,Vti1, 139 pmol protein, 77 nmol lipid; t-Sed5/Bos1,Sec22, 403 pmol protein, 92 nmol lipid. Addition of 5 μ l of v-SNAREs gave the following amounts of protein and lipid: v-Snc2, 85 pmol protein, 5.4 nmol lipid; v-Nyv1, 45 pmol protein, 5 nmol lipid; v-Bet1, 87 pmol protein, 5.6 nmol lipid. The plate was transferred to a 37 °C fluorescent plate reader and NBD fluorescence (excitation 460 nm, emission 538 nm) monitored at 2-min intervals. At 120 min, detergent was added to determine NBD fluorescence at infinite dilution. Data are expressed as rounds of fusion measured as fold lipid dilution^{5,17}.

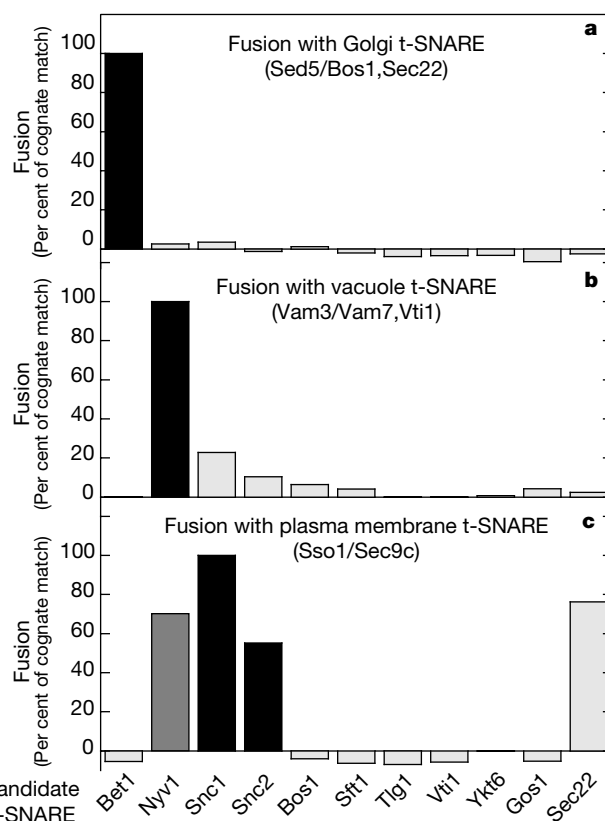


Figure 4 Specificity of the three functional t-SNARE complexes versus all other potential v-SNAREs. Liposomes containing the ER–Golgi t-SNARE complex (t-Sed5/Bos1,Sec22; **a**), the vacuolar t-SNARE complex (t-Vam3/Vam7,Vti1; **b**), and plasma membrane t-SNARE complex (t-Sso1/Sec9c; **c**) were independently reconstituted in liposomes. These t-SNARE liposomes were fused with all 11 non-syntaxin SNAREs in the yeast genome. The extent of fusion at 120 min was normalized to the amount of cognate v-SNARE fusion (black bars). The extent of an inhibited fusion reaction (by a v-SNARE cytoplasmic domain) was subtracted as a background for **a** (0.095 rounds) and **b** (0.063 rounds). A protein-free liposome background was subtracted from **c** (0.138 rounds). The maximal extent of fusion (100%) for the respective cognate fusion reactions was 1.85 (t-Sso1/Sec9 with v-Snc1), 1.0 (t-Vam3/Vam7,Vti1 with v-Nyv1), and 0.87 rounds of fusion (t-Sed5/Bos1,Sec22 with v-Bet1). Concentrations of protein and lipid for all the liposomes used here are shown in Fig. 1 and Table 2. Ykt6 was anchored to the membrane by the attachment of a synthetic C₂₀ isoprenoid¹³.

Table 1 Characterization of reconstituted liposomes used in Fig. 3

t-SNARE	v-SNARE	Localization	M_r	Moles of phospholipid per mole v- or t-SNARE	Copies of v- or t-SNARE per 45-nm vesicle*
t-Sed5/Bos1, Sec22		Golgi	118K	230	72
t-Vam3/Vam7, Vti1		Vacuole	99K	436	38
t-Sso1/Sec9c		PM	90K	250	65
	v-Bet1	ER/Golgi	16K	64	257
	v-Nyv1	Vacuole	53K	110	150
	v-Snc2	Golgi/PM	14K	64	255

ER, endoplasmic reticulum; PM, plasma membrane; Vac, vacuole.

* This calculation assumes a 45-nm diameter, that is a 225 Å outer leaflet radii and a 185 Å inner leaflet radii and a bilayer thickness of 40 Å, the surface area ($4\pi r^2$) of the outer leaflet is 636,000 Å² and the inner leaflet is 430,000 Å². These surface areas give a total phospholipid count of 16,400 molecules assuming an average phospholipid head group diameter of 65 Å², typically 60–80% of the SNAREs face outward (sensitive to thrombin or other protease digestion). Protein determinations were done as described²⁰ and lipid concentrations were measured by tracer tritiated DPPC in the lipid mixture.

that of a cognate match (black bars in Fig. 4). Indistinguishable results were obtained from the rates of fusion (data not shown).

Notably, only one additional combination resulted in bilayer fusion: Sec22 with the plasma membrane t-SNARE. The Golgi (Fig. 4a) and the vacuole (Fig. 4b) t-SNAREs only fused significantly with their respective cognate matches.

Although it is highly restricted in its potential to fuse (Fig. 4c), the plasma membrane t-SNARE is not strictly specific, unlike the Golgi or the vacuole t-SNAREs appear to be. All four of the v-SNAREs capable of fusing vesicles via the plasma membrane t-SNARE, Snc1, Snc2, Nyv1 and Sec22, have previously been identified as 'R' SNAREs³⁶ by virtue of a conserved arginine residue in the approximate centre of the predicted four-helix bundle^{10,36}. Because the conserved R can be mutated without significantly compromising secretion⁷, however, it now seems unlikely that this arginine is important in the fusion mechanism or in the specificity of SNARE pairing between membranes.

There is a fifth 'R' SNARE in yeast, Ykt6. When Ykt6 is lipid anchored, as it is in the cell, it cannot fuse (Fig. 4); however, an artificially protein-anchored version of Ykt6, like the other 'R' SNAREs (which are naturally protein anchored), does trigger fusion with the plasma membrane t-SNARE (Fig. 5). Evidently, lipid-anchoring of Ykt6 restricts it from acting as a v-SNARE, thereby indirectly increasing the effective specificity of the plasma membrane t-SNARE.

Discussion

The SNARE hypothesis. The concept that compartmental specificity in membrane fusion is encoded in SNARE proteins arose when the principle of cognate pairing was discovered in the context of regulated exocytosis¹. The essence of this 'SNARE hypothesis' was the deduction that "if there were no other source of specificity, only when complementary v-SNARE and t-SNARE pairs engage would a productive fusion event be initiated"¹. It was therefore proposed that vesicles dock specifically to target membranes when cognate

SNAREs link-up, and that the assembly of these SNARE complexes initiates the mechanism of bilayer fusion, which at that time was also thought to involve the associated proteins NSF (*N*-ethylmaleimide-sensitive) and SNAP (soluble NSF attachment protein). (The name SNARE derives from SNAP receptor.)

The role of the ATPase NSF and its adapter protein SNAP is now more precisely understood. Although these proteins are required to sustain continual fusion, they are not required for bilayer fusion per se³⁷. More specifically, NSF uses the energy from ATP hydrolysis to separate v- and t-SNAREs^{38,39}, dissociating the *cis*-SNARE complexes⁴⁰ that result from *trans*-SNARE complexes after fusion but not the *trans*-SNARE complexes actually mediating fusion⁴¹.

This suggested that the fusion mechanism might be even simpler than first thought, involving only the SNARE proteins. This view was made even more plausible by the pin-like topology predicted for *trans*-SNARE complexes¹¹. It is now clear that the assembly of SNARE complexes between membranes not only initiates the process of bilayer fusion, it is the mechanism of fusion². This, in turn, has made it possible for us forcefully and directly to establish the extent to which the pattern of membrane fusion processes in cells is encoded in the intrinsic physico-chemical properties of the SNARE proteins themselves.

The specificity of intracellular membrane fusion can, in principle, be encoded in SNARE proteins according to any one or a combination of three distinct mechanisms: (1) by restrictions on their ability to encounter each other owing to their differential distribution within cells. These patterns are encoded in their topogenic signals⁴². As these signals can only be read out in cells they cannot affect the specificity of fusion of bilayers by the isolated SNAREs. (2) By the pattern of interaction of SNARE proteins with regulatory proteins that could restrict their activity spatially or control it temporally. Because other proteins are involved, this mechanism also cannot operate when SNAREs are isolated. (3) By their inherent capacity (or incapacity) to pair between bilayers and then promote fusion in the relevant time frame. As these physico-chemical properties are

Table 2 Characterization of reconstituted liposomes used in Fig. 4

t-SNARE	Candidate v-SNARE	Localization	M_r	Moles of phospholipid per mole v- or t-SNARE	Copies of v- or t-SNARE per 45-nm vesicle*
t-Sed5/Bos1, Sec22		Golgi	118K	361	45
t-Vam3/Vam7, Vti1		Vacuole	99K	1,177	14
t-Sso1/Sec9c		PM	90K	633	26
	v-Bet1	ER/Golgi	16K	72	229
	v-Nyv1	Vacuole	53K	93	177
	v-Snc1	Golgi/PM	14K	49	335
	v-Snc2	Golgi/PM	14K	159	103
	Bos1	ER/Golgi	28K	63	261
	Sft1	Golgi	37K	15	1,111
	Tlg1	Golgi/E	26K	74	222
	Vti1	Vac/Golgi/E	28K	97	169
	Ykt6†	ER/Golgi	23K	110	149
	Gos1	Golgi	26K	33	490
	Sec22	ER/Golgi	27K	62	264

E, endosome; ER, endoplasmic reticulum; PM, plasma membrane. Candidate v-SNAREs known to function as v-SNAREs cognate to a t-SNARE being tested are indicated by the prefix 'v'.

* This calculation assumes a 45-nm diameter as in Table 1. Typically 60–80% of the SNAREs face outward (sensitive to thrombin or other protease digestion).

† Lipid-anchored with maleimido-propionic acid geranylgeranyl ester (M-C20)¹³.

retained in isolation, they will determine the pattern of specificity of bilayer fusion mediated by isolated SNARE proteins, as measured here.

Previous studies provided strong evidence for the SNARE hypothesis (for reviews, see refs 4, 28, 43–45). SNARE proteins are distributed in highly specific patterns among subcellular compartments that closely reflect the patterns of protein traffic between them. The same group of SNAREs whose localization is linked to compartments engaging in fusion can usually be isolated in stable molecular complexes (defining ‘cognate’ SNAREs) and is required for fusion processes involving these compartments *in vivo* and/or cell-free systems.

Specificity of fusion mediated by isolated SNARE proteins. To a marked degree, the pattern of transport pathways among the ER, Golgi apparatus, plasma membrane and vacuole are faithfully recapitulated by isolated SNARE proteins drawn from these compartments (Figs 3 and 4). We systematically tested all of the membrane-anchored, non-syntaxin SNARE proteins encoded in the yeast genome—each a potential candidate for function as a v-SNARE—for their capacity to mediate fusion with vesicles bearing t-SNAREs that represent the plasma membrane, the Golgi and the vacuole (Fig. 4). Of the total of 33 possible combinations, only the 4 previously known cognate combinations and 2 others (Nyv1 and Sec22; both in combination with the plasma membrane t-SNARE) were found to mediate fusion reactions. The Golgi and vacuole t-SNAREs each only mediated fusion with a single SNARE—the cognate v-SNARE. At present, we do not know whether those combinations that do not fuse may unproductively dock vesicles. Where this has been studied, they do not dock appreciably⁸.

Fusion mediated by pairing between the vacuolar v-SNARE and the plasma membrane t-SNARE, if it occurred in yeast cells, would correspond to the pathway of lysosomal secretion, which occurs in many types of animal cells and in lower eukaryotes⁴⁶. This suggests that fusion mediated by the Nyv1/plasma membrane pairing is probably a physiologically relevant cognate match. At the very least, this fusogenic combination cannot be considered non-cognate with any confidence.

Thus, Sec22 and the plasma membrane t-SNARE form the only known non-cognate combination that is fusogenic. Sec22 functions as a light chain in the Golgi t-SNARE Sed5/Bos1, Sec22 (ref. 8) and forms *cis*-SNARE complexes in the Golgi, which would be expected to limit the availability of Sec22 free to react with plasma membrane t-SNAREs *in vivo*. According to this model, the specificity of Sec22 would result in part from its inherent lack of reactivity as a v-SNARE

towards t-SNAREs other than that of the plasma membrane (Fig. 4), and in part from its interactions with other SNAREs that make Sec22 less available as a v-SNARE for the plasma membrane. All of these properties are, of course, inherent to the SNARE proteins themselves (mechanism (3) above). It is also likely that localization to the Golgi (mechanism (1)) and local activation (mechanism (2)) help to buttress the incomplete specificity of this SNARE.

Ykt6 is interesting because its cytoplasmic domain is capable of functioning as a v-SNARE when it is artificially anchored by a membrane-spanning peptide segment, but not when it is anchored by its native C₂₀ prenyl group²⁶. Evidently, the cell has used a lipid anchor as a device to prevent Ykt6 from functioning as a v-SNARE (Fig. 5), while still allowing it the potential to function as a t-SNARE light chain, possibly in combination with the syntaxins Sed5 or Vam3. All of this is encoded in the *YKT6* gene and retained in the physical chemistry of the isolated SNARE protein.

Two studies^{47,48} that used soluble SNAREs to assess specificity came to a very different conclusion. Their results indicated that there may be no inherent specificity in SNARE partnering in fusion. However, the current work vividly illustrates that membrane anchors as well as a direct functional read-out of bilayer fusion are needed to assess specificity. Together, the evidence suggests that SNARE pairing between bilayers may be inherently more specific than artificial pairing in solution. This is supported by the principle of topological restriction of bilayer fusion⁸, where the same four SNAREs will or will not assemble depending on precisely how their membrane anchors are distributed between the two bilayers, even though the same set of cytoplasmic domains are present in every case.

Indeed, the pattern of binding of soluble SNAREs should be a poor predictor of the capacity of membrane-anchored SNAREs to engage, because bilayer fusion results from a coupled process in which some of the energy made available from the assembly of SNAREpins is used to do work on the bilayers¹³, correspondingly reducing the energy stabilizing the SNAREpins. The energy thereby stored in the lipid bilayers is released when fusion occurs as *trans*-SNARE complexes concomitantly relax to *cis*-SNARE complexes^{2,10,11}. It follows from these basic energetic considerations that a (non-cognate) pairing that can occur in solution will not occur between bilayers if the energy stabilizing it in solution, however favourable in itself, is insufficient to do the requisite work on the bilayers when these processes are coupled by membrane anchoring. Thus, the coupling of SNARE assembly to do work on the bilayers enhances selectivity and also meets a basic requirement for fusion.

Conclusions and perspectives

The data now available indicate that, to a very high degree, the specificity of intracellular membrane fusion is encoded in the inherent physico-chemical properties of their SNARE proteins. As a result, the pattern of membrane flow among the compartments of a cell is, almost without exception, faithfully recapitulated by constituent SNARE proteins in isolation from the cell, exactly as predicted by the SNARE hypothesis¹.

It is likely that the specificity of membrane flow—although already greatly restricted by the level of specificity inherent in their SNARE proteins (Fig. 4)—is nonetheless buttressed by patterns of localization and activation of the SNAREs by regulatory proteins, including tethering and Rab GTPase proteins. Various studies (reviewed in refs 43–45, 49) indicate that vesicles can be initially captured at designated portions of target membranes by long, probably flexible tethers extending from the membranes. Following tethering, *trans*-SNARE complexes assemble to dock and fuse the vesicles. Different targets can use distinct sets of tethering and vesicle-bound Rab proteins, which are active only in their GTP-bound state.

SNAREs and regulatory systems, including Rabs and tethers

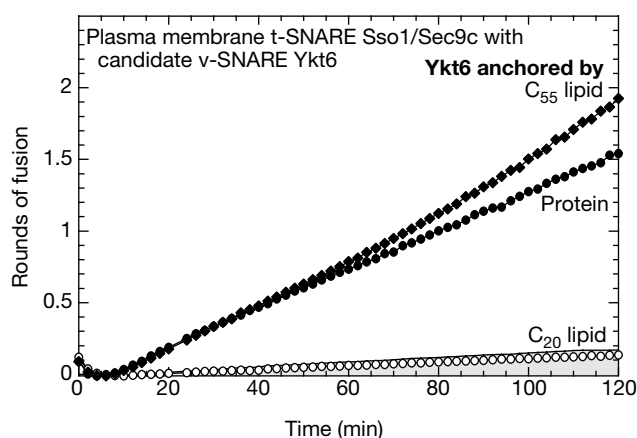


Figure 5 Ykt6 anchored with a short-chain isoprenoid does not fuse. Liposomes containing Ykt6 anchored with a C₂₀ isoprenoid (similar to its endogenous lipid) or a C₅₅ isoprenoid was compared with a chimaeric Ykt6 with a traditional transmembrane domain. The grey-shaded area corresponds to the extent of fusion of an inhibited fusion reaction with soluble Snc2.

probably have complementary rather than redundant properties. Tethers and Rabs respectively permit fine spatial and temporal control of SNARE-mediated fusion, harnessing this engine to position membrane flow dynamically for cell growth, movement and development, and controlling its rate according to the GTP-charge of Rab proteins. By contrast, SNAREs provide the core, evolutionarily conserved and usually static code for the specificity of membrane compartments as a whole (that is, apical versus basal plasma membrane versus Golgi versus ER versus endosome, and so on). This code is inextricably linked to the physical core of intracellular membrane fusion, the SNAREpin, which must be assembled from two cognate parts (v-SNARE and t-SNARE) that reside in a different membranes. □

Methods

DNA manipulation and plasmid construction

Standard genetic manipulations were performed throughout. All polymerase chain reaction (PCR) procedures were done with *Pwo* polymerase (BM). All other DNA modifying enzymes were from New England Biolabs. The *E. coli* strain DH5 α (Gibco) was used for standard cloning.

Plasmid pJM84 (untagged Sso1) was generated by PCR using *S. cerevisiae* genomic DNA as template. An *EcoRI*–*XhoI* fragment was ligated into pET24 (Novagen). An octahistidine tag was added to the amino terminus of the Sso1 gene by inserting a double-stranded oligo into the *Bam*HI–*EcoRI* sites of pJM84 resulting in pJM88 (His₈–Sso1). Plasmid pJM81 (Snc2–His₆) was produced by introduction of a 366-base-pair PCR fragment containing the *SNC2* open reading frame amplified from *S. cerevisiae* genomic DNA and ligated into the *NcoI*–*XhoI* sites of pET28a. We expressed Snc1–His₆ from pJM91. We amplified *SNC1* from *S. cerevisiae* genomic DNA ligated into the *NcoI*–*XhoI* sites of pET28a. The SNAP25 portion of Sec9 (Sec9c) and a truncated Snc2 (without a transmembrane domain) were expressed as GST fusion proteins from the plasmids BB442 and BB464a, respectively²³. The coding sequence of *TLG1* was amplified by PCR from *S. cerevisiae* genomic DNA and ligated in the *NdeI* and *Bam*HI sites of pET28a (pJM124). Four YKT6 plasmids were derived from pGEX–Ykt6⁹. Cysteine 76 was replaced with serine by site-directed mutagenesis to give pJM127. We added a two amino-acid linker (PR) was added between the thrombin cleavage site and Ykt6 to increase the efficiency of thrombin cleavage to give pJM128, which expresses GST–PR–Ykt6 (C76S). We made a chimaeric Ykt6 containing the transmembrane domain of Gos1 (pJM125) by ligating a PCR fragment containing Ykt6/Gos1TMD²⁶ into pGEX–Ykt6.

Protein expression

We produced Snc2–His₆ by expression of pJM81 in the BL21 (DE3) *E. coli* strain. Protein expression was induced with 0.2 mM IPTG for 4 h at 37 °C. The cells were collected by centrifugation and resuspended in breaking buffer (25 mM HEPES–KOH, pH 7.7, 400 mM KCl, 10% (w/v) glycerol, 2 mM β -mercaptoethanol and 1 EDTA-free protease inhibitor tablet and 4% Triton X-100). We disrupted the cells by two passages through an Emulsiflex C5 cell disrupter (Avestin). Cell debris was removed by centrifugation, and the supernatant was added to Ni-NTA beads (Qiagen) and bound overnight. We washed the beads with TX-100 wash buffer (25 mM HEPES–KOH, pH 7.7, 400 mM KCl, 10% (w/v) glycerol, 2 mM β -mercaptoethanol, 1% Triton X-100), and exchanged Triton X-100 for octyl- β -D-glucopyranoside (OG) by washing the column in OG wash buffer (25 mM HEPES–KOH, pH 7.7, 100 mM KCl, 10% (w/v) glycerol, 2 mM β -mercaptoethanol, 50 mM imidazole–OAc, pH 7.5, 1.0% (w/v) OG). The protein was eluted with a linear, seven-column-volume imidazole gradient from 50 mM to 500 mM. Peak fractions were pooled and frozen at –80 °C. All other His-tagged proteins were similarly produced (Snc1–His₆, His₈–Sso1, His₆–Tlg1), with the following exception: expression of His₈–Sso1 (in the Epicurian Coli BL21–Codon Plus (DE3)–RIL (Stratagene) *E. coli* strain) was induced at an absorbance at 600 nm of ~0.7, with 0.2 mM IPTG for 16 h at 16 °C.

We expressed GST–Sec9c from BB442 in Epicurian Coli BL21–Codon Plus (DE3)–RIL (Stratagene) *E. coli* strain. The culture was induced with 0.2 mM IPTG for 2 h at 25 °C. A short (2-h) induction was imperative to reduce the amount of *in vivo* proteolysis of GST–Sec9c. We collected the cells by centrifugation and resuspended them in breaking buffer (25 mM HEPES–KOH, pH 7.7, 400 mM KCl, 10% (w/v) glycerol, 2 mM β -mercaptoethanol, and 1 complete protease inhibitor Tablet). Cell debris was removed by centrifugation and bound in batch to glutathione agarose bead (Sigma) equilibrated in breaking buffer overnight at 4 °C. The GSH-beads were washed with breaking buffer and eluted in batch by adding 2.0 ml of breaking buffer containing 10 mM reduced glutathione, followed by a 1-ml elution. We expressed GST–Snc2 Δ TMD from BB464a in BL21 (DE3) *E. coli* and purified it as for GST–Sec9c, except that the protein was produced at 37 °C. We expressed and purified GST–Gos1, GST–Sft1 and GST–Ykt6 as described³⁰, except that GST–Sft1 was reconstituted as a GST fusion protein and cleaved in liposomes. We produced SNARE proteins involved in homotypic vacuolar fusion (Vam3, Vam7, Vti1 and Nyv1) as described⁷, and SNARE proteins involved in ER–Golgi fusion (Sed5, Bos1, Sec22 and Bet1) as described⁸.

Reconstitution

All of the yeast t-SNARE complexes were reconstituted from individually expressed proteins. The plasma membrane t-SNARE complex (Sso1/Sec9c) was formed by mixing

His₈–Sso1 and GST–Sec9c in ~1:5 molar ratio in a 1.5-ml microfuge tube. As GST–Sec9c does not contain detergent, OG was added (13 μ l from a 20% (w/w) stock) to 1% OG final concentration in a total of 500 μ l. Complex formation proceeded for ~24 h at 4 °C on a rotating wheel. The ER–Golgi t-SNARE complex and the vacuolar t-SNARE complex were similarly formed by using described procedures^{8,9}.

We used each of the t-SNARE complexes (500 μ l) to resuspend a lipid film containing 1,500 nmole POPC:DOPS in an 85:15 mole ratio. We carried out liposome formation by detergent dilution and dialysis, and liposome isolation by flotation as described^{8,9}. We removed GST from GST–Sec9 and GST–Sed5 by cleavage with thrombin (0.05–0.1 units of thrombin per microlitre of liposome) for 2 h at room temperature. Thrombin activity was inhibited by the addition of AEBSF (2 mM final concentration). Reconstitution of Snc2 into donor liposomes was done as described for VAMP¹⁷. Where indicated, we used less than 100 μ l of Snc2 protein in a given reconstitution. In these cases, the volume was made up to 100 μ l with buffer A₄₀₀ (25 mM HEPES–KOH, 400 mM KCl, 10% (w/v) glycerol, 1 mM dithiothreitol, pH 7.4) containing 1% (w/v) OG. We reconstituted Bet1 and GST–Nyv1 as described^{8,9}, except that 150 μ l of GST–Nyv1 liposomes were collected instead of 75 μ l. Ykt6 was lipid-anchored in detergent solution and reconstituted by detergent dilution and dialysis as described for the neuronal t-SNARE complex¹³.

Fusion assay

We carried out standard fusion assays as described² with modifications¹⁷. We mixed 45 μ l of unlabelled t-SNARE liposomes on ice with 5 μ l of fluorescently labelled v-SNARE liposomes in a 96-well Fluoronunc polysorp plate (Nunc). The plate was immediately placed in a 37 °C fluorescent plate reader (Floroskan II, Labsystems) without any prior pre-incubation, and NDB fluorescence was measured (excitation 460 nm, emission 538 nm) at 2-min intervals. At the end of 120 min, 10 μ l of 2.5% (w/v) *n*-dodecylmaltoside (Boehringer Mannheim) was added to determine NBD fluorescence at infinite dilution. We first converted the kinetic fusion data to per-cent detergent signal as described¹⁷, and then converted it to rounds of fusion using the equation $y = (0.49666 \times e^{(0.036031x)} - (0.50597 \times e^{(-0.053946x)}))$ where y is the fold lipid dilution and x is the per-cent dodecylmaltoside signal at a given time interval, as described^{5,17}.

Received 30 May; accepted 18 July 2000.

- Söllner, T. *et al.* SNAP receptors implicated in vesicle targeting and fusion. *Nature* **362**, 318–324 (1993).
- Weber, T. *et al.* SNAREpins: minimal machinery for membrane fusion. *Cell* **92**, 759–772 (1998).
- Rothman, J. E. Mechanisms of intracellular protein transport. *Nature* **372**, 55–63 (1994).
- Pelham, H. R. SNAREs and the secretory pathway—lessons from yeast. *Exp. Cell Res.* **247**, 1–8 (1999).
- Parlati, F. *et al.* Rapid and efficient fusion of phospholipid vesicles by the alpha-helical core of a SNARE complex in the absence of an N-terminal regulatory domain. *Proc. Natl Acad. Sci. USA* **96**, 12565–12570 (1999).
- Nickel, W. *et al.* Content mixing and membrane integrity during membrane fusion driven by pairing of isolated v-SNAREs and t-SNAREs. *Proc. Natl Acad. Sci. USA* **96**, 12571–12576 (1999).
- Chen, Y. A., Scales, S. J., Patel, S. M., Doung, Y. C. & Scheller, R. H. SNARE complex formation is triggered by Ca²⁺ and drives membrane fusion. *Cell* **97**, 165–174 (1999).
- Parlati, F. *et al.* Topological restriction of SNARE-dependent membrane fusion. *Nature* **407**, 194–198 (2000).
- Fukuda, R. *et al.* Functional architecture of an intracellular t-SNARE. *Nature* **407**, 198–202 (2000).
- Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* **395**, 347–353 (1998).
- Hanson, P. L., Roth, R., Morisaki, H., Jahn, R. & Heuser, J. E. Structure and conformational changes in NSF and its membrane receptor complexes visualized by quick-freeze/deep-etch electron microscopy. *Cell* **90**, 523–535 (1997).
- Fasshauer, D., Otto, H., Eliason, W. K., Jahn, R. & Brunger, A. T. Structural changes are associated with soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor complex formation. *J. Biol. Chem.* **272**, 28036–28041 (1997).
- McNew, J. A. *et al.* Close is not enough: SNARE-dependent membrane fusion requires an active mechanism that transduces force to membrane anchors. *J. Cell Biol.* **150**, 105–118 (2000).
- Nicholson, K. L. *et al.* Regulation of SNARE complex assembly by an N-terminal domain of the t-SNARE Sso1p. *Nature Struct. Biol.* **5**, 793–802 (1998).
- Hua, S. Y. & Charlton, M. P. Activity-dependent changes in partial VAMP complexes during neurotransmitter release. *Nature Neurosci.* **2**, 1078–1083 (1999).
- Xu, T. *et al.* Inhibition of SNARE complex assembly differentially affects kinetic components of exocytosis. *Cell* **99**, 713–722 (1999).
- McNew, J. A., Weber, T., Engelman, D. M., Söllner, T. H. & Rothman, J. E. The length of the flexible SNAREpin juxtamembrane region is a critical determinant of SNARE-dependent fusion. *Mol. Cell* **4**, 415–421 (1999).
- Weimbs, T., Mostov, K., Low, S. H. & Hofmann, K. A model for structural similarity between different SNARE complexes based on sequence relationships. *Trends Cell. Biol.* **8**, 260–262 (1998).
- Brennwald, P. *et al.* Sec9 is a SNAP-25-like component of a yeast SNARE complex that may be the effector of Sec4 function in exocytosis. *Cell* **79**, 245–258 (1994).
- Aalto, M. K., Ronne, H. & Keranen, S. Yeast syntaxins Sso1p and Sso2p belong to a family of related membrane proteins that function in vesicular transport. *EMBO J.* **12**, 4095–4104 (1993).
- Protopopov, V., Govindan, B., Novick, P. & Gerst, J. E. Homologs of the synaptobrevin/VAMP family of synaptic vesicle proteins function on the late secretory pathway in *S. cerevisiae*. *Cell* **74**, 855–861 (1993).
- Gerst, J. E., Rodgers, L., Riggs, M. & Wigler, M. SNCL1, a yeast homolog of the synaptic vesicle-associated membrane protein/synaptobrevin gene family: genetic interactions with the RAS and CAP genes. *Proc. Natl Acad. Sci. USA* **89**, 4338–4342 (1992).
- Katz, L., Hanson, P. L., Heuser, J. E. & Brennwald, P. Genetic and morphological analyses reveal a critical interaction between the C-termini of two SNARE proteins and a parallel four helical arrangement for the exocytic SNARE complex. *EMBO J.* **17**, 6200–6209 (1998).
- Gerst, J. E. Conserved alpha-helical segments on yeast homologs of the synaptobrevin/VAMP family of v-SNAREs mediate exocytic function. *J. Biol. Chem.* **272**, 16591–16598 (1997).

25. Neiman, A. M. Prospore membrane formation defines a developmentally regulated branch of the secretory pathway in yeast. *J. Cell Biol.* **140**, 29–37 (1998).
26. McNew, J. A. *et al.* Ykt6p, a prenylated SNARE essential for endoplasmic reticulum-Golgi transport. *J. Biol. Chem.* **272**, 17776–17783 (1997).
27. Newman, A. P., Shim, J. & Ferro-Novick, S. BET1, BOS1, and SEC22 are members of a group of interacting yeast genes required for transport from the endoplasmic reticulum to the Golgi complex. *Mol. Cell. Biol.* **10**, 3405–3414 (1990).
28. Nichols, B. J. & Pelham, H. R. SNAREs and membrane fusion in the Golgi apparatus. *Biochim. Biophys. Acta* **1404**, 9–31 (1998).
29. Banfield, D. K., Lewis, M. J. & Pelham, H. R. A SNARE-like protein required for traffic through the Golgi complex. *Nature* **375**, 806–809 (1995).
30. McNew, J. A. *et al.* Gos1p, a *Saccharomyces cerevisiae* SNARE protein involved in Golgi transport. *FEBS Lett.* **435**, 89–95 (1998).
31. Søgaard, M. *et al.* A rab protein is required for the assembly of SNARE complexes in the docking of transport vesicles. *Cell* **78**, 937–948 (1994).
32. Lupashin, V. V., Pokrovskaya, I. D., McNew, J. A. & Waters, M. G. Characterization of a novel yeast SNARE protein implicated in Golgi retrograde traffic. *Mol. Biol. Cell* **8**, 2659–2676 (1997).
33. von Mollard, G. F., Nothwehr, S. F. & Stevens, T. H. The yeast v-SNARE Vti1p mediates two vesicle transport pathways through interactions with the t-SNAREs Sed5p and Pep12p. *J. Cell Biol.* **137**, 1511–1524 (1997).
34. Abeliovich, H., Grote, E., Novick, P. & Ferro-Novick, S. Tlg2p, a yeast syntaxin homolog that resides on the Golgi and endocytic structures. *J. Biol. Chem.* **273**, 11719–11727 (1998).
35. Fischer von Mollard, G. & Stevens, T. H. The *Saccharomyces cerevisiae* v-SNARE Vti1p is required for multiple membrane transport pathways to the vacuole. *Mol. Biol. Cell* **10**, 1719–1732 (1999).
36. Fasshauer, D., Sutton, R. B., Brunger, A. T. & Jahn, R. Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *Proc. Natl Acad. Sci. USA* **95**, 15781–15786 (1998).
37. Mayer, A., Wickner, W. & Haas, A. Sec18p (NSF)-driven release of Sec17p (alpha-SNAP) can precede docking and fusion of yeast vacuoles. *Cell* **85**, 83–94 (1996).
38. Ungermann, C., Nichols, B. J., Pelham, H. R. & Wickner, W. A vacuolar v-t-SNARE complex, the predominant form *in vivo* and on isolated vacuoles, is disassembled and activated for docking and fusion. *J. Cell Biol.* **140**, 61–69 (1998).
39. Söllner, T., Bennett, M. K., Whiteheart, S. W., Scheller, R. H. & Rothman, J. E. A protein assembly-disassembly pathway *in vitro* that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. *Cell* **75**, 409–418 (1993).
40. Nichols, B. J., Ungermann, C., Pelham, H. R., Wickner, W. T. & Haas, A. Homotypic vacuolar fusion mediated by t- and v-SNAREs. *Nature* **387**, 199–202 (1997).
41. Weber, T. *et al.* SNAREpins are functionally resistant to disruption by NSF and α -SNAP. *J. Cell Biol.* **149**, 1063–1072 (2000).
42. Rayner, J. C. & Pelham, H. R. Transmembrane domain-dependent sorting of proteins to the ER and plasma membrane in yeast. *EMBO J.* **16**, 1832–1841 (1997).
43. Waters, M. G. & Pfeffer, S. R. Membrane tethering in intracellular transport. *Curr. Opin. Cell Biol.* **11**, 453–459 (1999).
44. Gonzalez, L. Jr. & Scheller, R. H. Regulation of membrane trafficking: structural insights from a Rab/effector complex. *Cell* **96**, 755–758 (1999).
45. Mellman, I. & Warren, G. The road taken: past and future foundations of membrane traffic. *Cell* **100**, 99–112 (2000).
46. Stinchcombe, J. C. & Griffiths, G. M. Regulated secretion from hemopoietic cells. *J. Cell Biol.* **147**, 1–6 (1999).
47. Fasshauer, D., Antonin, W., Margittai, M., Pabst, S. & Jahn, R. Mixed and non-cognate SNARE complexes. Characterization of assembly and biophysical properties. *J. Biol. Chem.* **274**, 15440–15446 (1999).
48. Yang, B. *et al.* SNARE interactions are not selective. Implications for membrane fusion specificity. *J. Biol. Chem.* **274**, 5649–5653 (1999).
49. Novick, P. & Zerial, M. The diversity of Rab proteins in vesicle transport. *Curr. Opin. Cell Biol.* **9**, 496–504 (1997).
50. Schaffner, W. & Weissmann, C. A rapid, sensitive, and specific method for the determination of protein in dilute solution. *Anal. Biochem.* **56**, 502–514 (1973).

Acknowledgements

We wish to thank Dr Paul A. Marks for two decades of exemplary leadership of Memorial Sloan-Kettering Cancer Center and for creating the environment that made possible the discovery of SNARE proteins and their role in cell biology. We also thank L. Katz and P. Brennwald for plasmids, B. Brügger for help in the construction of the TLG1 clone, and T. Melia for cryo-electron microscopy of liposomes. Research was supported by an NIH grant (to J.E.R.) and postdoctoral fellowships of the Medical Research Council of Canada (F.P.), the NIH (J.M.), and the Japanese Society for the Promotion of Science (R.F.).

Correspondence and requests for materials should be addressed to J.E.R. (e-mail: j-rothman@ski.mskcc.org).

Laboratory detection of X-ray fringes with a grazing-incidence interferometer

Webster Cash*, Ann Shipley*, Steve Osterman* & Marshall Joy†

* University of Colorado, Boulder, Colorado 80309-0389, USA

† NASA/Marshall Space Flight Center, Huntsville, Alabama 35812, USA

Starting with Galileo's observations of the Solar System, improvements of an order of magnitude in either the sensitivity or resolution of astronomical instruments have always brought revolutionary discoveries. The X-ray band of the spectrum, where exotic objects can have extremely high surface brightness¹, is ideally suited for significant improvements in imaging, but progress has been impeded by a lack of optics of sufficiently high sensitivity and quality. Here we present an X-ray interferometer design that is practical for adaptation to astronomical observatories. Our prototype interferometer, having just under one millimetre of baseline, creates fringes at 1.25 keV with an angular resolution of 100 milliarcseconds. With a larger version in orbit it will be possible to resolve X-ray sources at 10^{-7} arcseconds, three orders of magnitude better than the finest-resolution images ever achieved on the sky (in the radio part of the spectrum) and over one million times better than the current best X-ray images. With such resolutions, we can study the environments of pulsars, resolve and then model relativistic blast waves, image material falling into a black hole, watch the physical formation of astrophysical jets, and study the dynamos of stellar coronae.

The Chandra X-ray Observatory² now provides sub-arcsecond angular resolution, a superlative achievement for a grazing incidence telescope, but far short of the X-ray diffraction limit. X-ray interferometers have been made from Laue crystals³ since 1965, but have found limited application in astronomy owing to low efficiency. Frank⁴ suggested splitting the amplitude of an X-ray beam using a thin metal foil but the system has never found application. In 1932, Kellstrom used a Lloyd's Mirror geometry and a Fresnel bent mirror to create X-ray fringes⁵. However, the Lloyd's geometry is poorly suited for astronomy as it requires the mirror to reflect at a tiny graze angle (that is, below one arcminute).

We have recently realized that there is a highly practical means of mixing X-ray wavefronts without a beam splitter. The principle is the same as that for the Fresnel bent-mirror geometry⁶, but is particularly suitable for astronomy at grazing incidence. As shown in Fig. 1, a wavefront impinging on two flat mirrors. The beams are

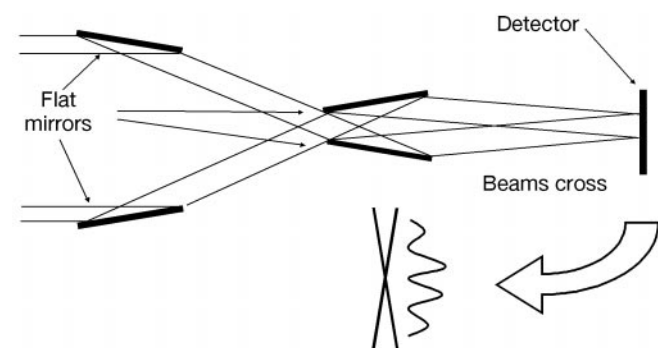


Figure 1 Four flat mirrors at grazing incidence can create a practical X-ray interferometer. A beam mixer is made from the two grazing flats. As the angle between the directions of the converging beams drops, the spacing of the fringes on the detector rises.

steered onto a second pair of flat mirrors for redirection into quasi-parallel convergence. The beam combiner must cause the wavefronts to cross at the detector at a very low relative angle to create a large fringe amplification. The wavelength of the fringes on the detector is given by $\lambda L/d$, where d is the separation of the secondary mirrors and L is the distance from the mirrors to the detector where the beams cross. If L/d is sufficiently large, the fringes become macroscopic. For example if L/d has a value of one million then 1 nm radiation will create fringes with 1 mm spacing. If the baseline, d , is 1 mm, then L needs to be 1 km, whereas if d is 30 cm, sufficient to build an observatory, then L rises to 300 km. This requires the detector of the interferometer to reside on a separate spacecraft, flying in formation—a task that is challenging, but achievable with today's spacecraft technology. Alternatively, if L/d is small, then tiny fringes can be created⁷.

Scattering and absorption properties of X-rays preclude the use of many standard techniques of visible light interferometry, and stringent tolerances apply to the optical components. However, a beam that reflects off a surface at grazing incidence experiences a reduction in the phase shift from a surface error. If the mirror has a hill, valley or displacement of size δ , then at normal incidence, the wavefront would experience a shift of 2δ , but at a graze angle of θ , the error would be only $2\delta\sin\theta$. Thus, if the phase emerging at a given point is not to change by more than $\lambda_X/10$, the surface displacement must not exceed $\lambda_X/(20\sin\theta)$, where λ_X is the wavelength of the X-ray and θ is the graze angle. To reflect 6-keV X-rays efficiently requires a 0.5-degree graze angle ($\sin\theta \approx 0.01$), leading to a full factor of 100 reduction in the required stability of the optic. In fact, the grazing incidence optic can move five times the wavelength of the X-ray before the fringe shifts by one tenth, so the typical position tolerance is about 1.5 nm instead of 15 pm—a great advantage. But this is still a tight tolerance. A mirror with surface errors of about 1.5 nm is usually considered to be ' $\lambda/400$ ' (632.8 nm), where a typical good mirror purchased commercially is near $\lambda/20$. Figuring mirrors to $\lambda/400$ or better is possible, but near the limit of currently achievable quality. Furthermore, figuring the mirror to a precise asphere (or sphere) is even more difficult. The best mirrors available are simple flat mirrors, so a design that uses only flat mirrors is much easier to build.

We have built and tested an interferometer of this simple four-flat-mirror design. Tests were performed in a 120-m vacuum facility at the Marshall Space Flight Center. X-rays were generated in a Manson Model 5 electron impact source with a magnesium target and 2- μ m aluminium filter to create a beam that consists mostly of the Mg K line (1.25 keV). The beam passed through a 5- μ m exit slit and 16 m later entered the interferometer. An entrance aperture of two parallel slits ensured that the two primary mirrors were illuminated, but that none of the direct beam could enter the interferometer without reflecting off the optics. After passing the double slit, the wavefronts reflected on the primary mirrors. Each mirror was a 50 mm circular optical flat mirror set at 0.21 degrees to the incoming beam. Each was mounted in a precision manipulator that allowed fine rotational and translation adjustment from outside the vacuum tank. The centres of the primary mirrors were separated by approximately 0.5 mm. The primary mirrors created reflected wavefronts that crossed and were then reflected by the secondary mirror. The fronts of the secondary mirrors were 17 mm beyond the back of the primaries. These mirrors were also 50 mm in diameter, circular optical flat mirrors mounted on manipulators, and were separated by 0.5 mm at their centres. The wavefronts, when they emerged from the secondary mirrors, were very nearly parallel. They then travelled 100 m down the vacuum pipe to the detector, which was a charge-coupled device (CCD) with pixels 18 micrometres square.

When the system was turned on, each of the two beams created a

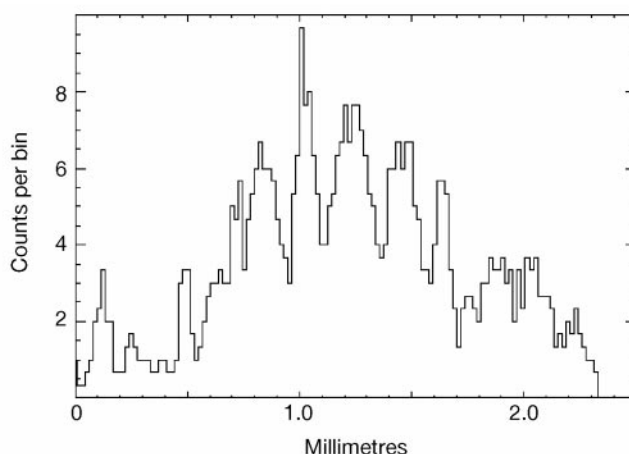


Figure 2 X-ray interference fringes were recorded in the laboratory. The 1.25-keV (Mg $K\alpha$) events were accumulated on a charge-coupled device (CCD) detector. The image

was rotated half a degree to align the fringes with the y -axis and then binned. A two-bin boxcar smoothing function was run over the histogram to suppress Poisson noise.

vertical stripe of illumination on the CCD about 1 mm (full-width zero-maximum) wide, which was expected from diffraction in the 0.2-mm mirror apertures. The final step was to fine-adjust the angles of the secondary mirrors so that the two stripes fell on top of each other at the CCD. The image was rotated about half a degree to allow for alignment to the vertical. The counts were then gathered into a histogram of events across the horizontal direction. A two-bin boxcar smoothing function was run across the data to suppress the Poisson noise. With $L = 100$ m, $d = 0.5$ mm and $\lambda = 1$ nm, we expect to see 0.2-mm fringes across the 1-mm stripe of illumination, exactly as seen in Fig. 2.

The fringes in Fig. 2, created on a 1-mm baseline, are sensitive to angular scales near one tenth of an arcsecond and open a practical pathway to ultra-high resolution. As the separation of the primary mirrors is increased, so is the angular resolution of the interferometer. Radio astronomers have been inverting interferometric fringes to create images for decades and the same is now being accomplished in the visible⁸.

In the X-ray band we have an advantage over visible light and radio interferometers. Our detectors record individual photons and tag each with an energy. A simple CCD gives resolution as good as $E/\delta E$ of 50 at 6 keV, while microcalorimeter devices approach a resolution of 1,000 at the same energy. With no loss of signal, data can be separated into separate sine waves of different frequencies. To the extent that the source does not change as a function of emission wavelength, this represents an improvement in our sampling of frequency space. Another advantage of interferometry can be seen in Table 1 where we show the fundamental tolerances which must be met as a function of the desired resolution.

Table 1 X-ray interferometric observatory requirements shown at four angular resolutions

Resolution (arcsec)	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Baseline (m)	1	10	100	1,000
Mirror size (cm)	3×100	3×100	3×100	3×100
Position stability (nm)	20	20	20	20
Angular stability (arcsec)	10^{-3}	10^{-4}	10^{-5}	10^{-6}
Angular stability (nm over baseline)	5	5	5	5
Figure	$\lambda/200$	$\lambda/200$	$\lambda/200$	$\lambda/200$
Polish ($\text{\AA}$, r.m.s.)	20	20	20	20
Angular knowledge (arcsec)	3×10^{-5}	3×10^{-6}	3×10^{-7}	3×10^{-8}
Angular knowledge (nm over baseline)	0.16	0.16	0.16	0.16
Position knowledge (nm)	2	2	2	2

The baseline grows from 1 m at 100 μs to 1 km at 100 nas. The quality and size of the optics do not change with system resolution. Only the angular stability and knowledge grow more constrained along with the resolution. However, the size of the position error that causes the pointing error does not grow, because the size of the interferometer does. A high-resolution interferometer has the same fabrication tolerances as a low-resolution system; however, it must hold the tolerance over a larger distance.

Once the tolerances have been met for an interferometer at any resolution, improvement can be achieved by simply increasing the baseline. The mirror quality does not need to improve, nor does the metrological precision—they must merely be held at the same level over a larger instrument.

Much of the work described here has been motivated by the desire to build functional X-ray interferometric observatories, as endorsed by the National Academy of Sciences⁹. The first observatory might feature a metre-class X-ray interferometer designed to achieve spatial resolution of 100 microarcseconds (μas) at 1 keV. The second mission would require a 100-m baseline in order to reach resolution as fine as 100 nanoarcseconds (nas) at 6 keV. Initial studies of the full missions, including spacecraft and launch requirements, have shown them to be technically challenging, but within the grasp of existing space technology. With 100- μas imaging we would be able to take pictures of the coronae of other stars. The nearest star, Alpha Centauri, is a binary, with one component the same size and temperature as our Sun. That disk, as viewed from Earth, is 7 milliarcseconds (mas) in diameter. Farther away, there are even more interesting targets. For example, at 50 pc lies AR Lacertae, a binary star with a period of just 1.98 days. X-ray profiles during eclipse indicate that much of the X-ray emission comes from plasma that is interconnected between the two stars¹⁰. Images will give us a probe of the physical state of plasma in a completely new astrophysical environment.

Resolution better than 1 μas puts within our technological grasp a basic goal of astronomy—to take a picture of a black hole. Such images would provide incontrovertible proof of existence of these objects. They would allow us to study the exotic physics at work in the immediate vicinity of black holes: the physics of the innermost accretion disk, the hard X-ray emitting corona, the formation of the relativistic jets, and the ‘plunging’ region in which material undergoes the final spiral through the black hole gravitational field towards the event horizon. $\square$

Received 8 May; accepted 27 June 2000.

- Holt, S. & McCray, R. Spectra of cosmic X-ray sources. *Annu. Rev. Astron. Astrophys.* **20**, 323–365 (1982).
- Weisskopf, M. C., O’Dell, S. L. & VanSpeybroeck, L. P. Advanced X-ray astrophysics facility (AXAF): An overview. *Proc. Soc. Photo-Opt. Eng.* **2805**, 2–7 (1996).
- Bonse, U. & Hart, M. An X-ray interferometer. *Appl. Phys. Lett.* **6**, 155–156 (1965).
- Franks, A. X-ray interferometers. United States Patent and Trademark Office, Washington DC, Patent Number 4,174,478, 1–7 (1979).
- Kellstrom, G. Experimentelle Untersuchungen über Interferenz- und Beugungserscheinungen bei Langwelligen Röntgenstrahlen. *Nova Acta Soc. Sci. Upsala* **8**, 5–66 (1932).
- Born, M. & Wolf, E. *Principles of Optics* 7th edn, 292–263 (Cambridge Univ. Press, Cambridge, 1999).
- Polack, F., Joyeux, D., Svatos, J. & Phalippou, D. Applications of wavefront division interferometers in soft x-rays. *Rev. Sci. Instrum.* **66**, 2180–2183 (1995).

8. Baldwin, J. E. *et al.* The first images from an optical aperture synthesis array. *Astron. Astrophys.* **306**, L13–L16 (1996).
9. McKee, C. & Taylor, J. (eds) *Astronomy and Astrophysics in the New Millennium* (National Academy Press, Washington DC, 2000).
10. Siarkowski, M., Pres, P., Drake, S. A., White, N. E. & Singh, K. P. The Coroneae of AR Lac II: The Spatial Structure. *Astrophys. J.* **473**, 470–482 (1996).

Acknowledgements

We thank N. White, D. Windt, J. Bixler, K. Doty, J. Carter, K. Herren, D. Goodman, J. Kolodziejczak, G. Zirnstein, T. Kester, M. Weisskopf and M. Karovska. This work was funded by NASA with additional technical support from the NASA Goddard Space Flight Center.

Correspondence and requests for materials should be addressed to W.C. (e-mail: cash@casa.colorado.edu).

Enhanced supercurrent density in polycrystalline $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ at 77 K from calcium doping of grain boundaries

G. Hammerl*, A. Schmehl*, R. R. Schulz*, B. Goetz*, H. Bielefeldt*, C. W. Schneider*, H. Hilgenkamp*† & J. Mannhart*

* Experimental Physics VI, Center for Electronic Correlations and Magnetism, Institute of Physics, University of Augsburg, D-86135 Augsburg, Germany

† Low Temperature Division and MESA* Institute, University of Twente, Box 217, 7500 AE Enschede, The Netherlands

With the discovery of high-temperature superconductivity¹, it seemed that the vision of superconducting power cables operating at the boiling point of liquid nitrogen (77 K) was close to realization. But it was soon found that the critical current density J_c of the supercurrents that can pass through these polycrystalline materials without destroying superconductivity is remarkably small^{1,2}. In many materials, J_c is suppressed at grain boundaries^{2–4}, by phenomena such as interface charging and bending of the electronic band structure^{5–9}. Partial replacement ('doping') of the yttrium in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ with calcium has been used to increase grain-boundary J_c values substantially, but only at temperatures much lower than 77 K (ref. 9). Here we show that preferentially overdoping the grain boundaries, relative to the grains themselves, yields values of J_c at 77 K that far exceed previously published values. Our results indicate that grain-boundary doping is a viable approach for producing a practical, cost-effective superconducting power cable operating at liquid-nitrogen temperatures.

The realization that the electronic properties of grain boundaries in high- T_c superconductors are controlled to a large extent by interface charging and band bending^{5–9} provided the clue to the enhancement of grain-boundary critical current densities by overdoping⁹ (that is, doping the grain boundaries beyond the composition that is optimal for superconductivity in the bulk). Overdoping reduces the built-in potential of the interfaces and increases the carrier density of the superconductors, so that the boundaries' tunnelling barriers are reduced in height and width. By partially replacing Y^{3+} by Ca^{2+} , Schmehl *et al.*⁹ experimentally investigated the effects of doping bicrystalline $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ films on the intergrain J_c . They found that calcium doping leads to strongly enhanced J_c values at low temperatures. At 77 K, however, the critical current densities were only insignificantly increased, because overdoping not only enhances the grain boundary coupling but also reduces the transition temperature T_c of the superconductors. This T_c reduction counteracts the positive effects the doping exerts on the

grain boundary coupling. To achieve enhanced coupling across the boundaries in combination with a large intragrain T_c value, we sought a way to overdope the superconductors locally at the grain boundaries, while keeping the grains optimally doped. By exploiting grain boundary diffusion, local doping of the boundaries has been tried before, using silver (Chaudhari, P. *et al.*, unpublished work), iron¹⁰ and platinum¹⁰ in experiments which were not designed to reduce charging or band bending. No J_c enhancements were observed in these studies.

To overdope the grain boundaries, we devised doping heterostructures, such as $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}/\text{Y}_{1-x}\text{Ca}_x\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ superlattices (Fig. 1a), anticipating that during film growth calcium would diffuse along the grain boundary into the $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ layers. As the diffusion coefficients in the grains are presumed to be small compared to coefficients for grain boundary diffusion, this process should enhance the Ca concentration at the $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ grain boundaries specifically. Therefore these layers were foreseen to have high intergrain J_c values at 77 K, combined with good superconducting intragrain properties. Thus, one material of the heterostructure, in this work $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$, was employed to optimize the properties of the bulk, whereas the other compound, $\text{Y}_{1-x}\text{Ca}_x\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ was used to improve the characteristics of the grain boundary.

To gain a detailed understanding of the possibilities and limitations

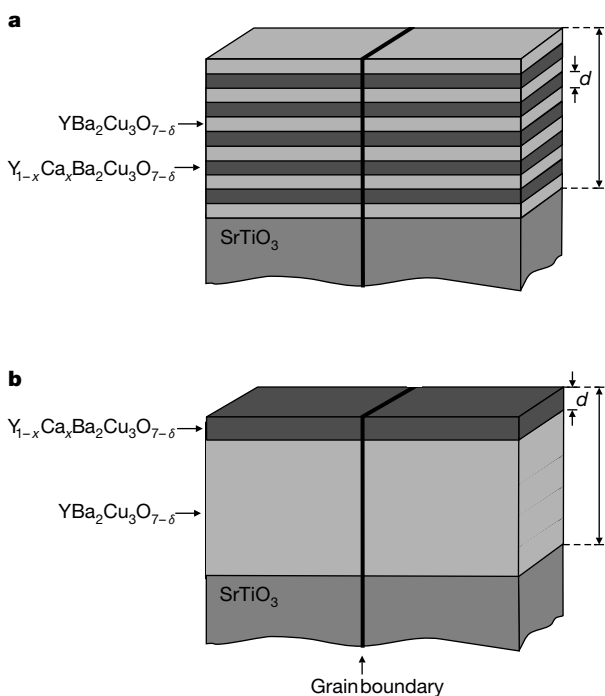


Figure 1 Illustration of the doping superlattices and bilayers investigated. Using standard deposition parameters⁹, the heterostructures and Ca-free control samples were grown within 10 minutes at a heater temperature of 760 °C by pulsed laser deposition on SrTiO_3 bicrystals containing symmetric $24 \pm 1^\circ$ [001] tilt grain boundaries. After deposition, the vacuum system was flooded with 0.4 bar of molecular oxygen, and the samples were cooled during 1 hour to 400 °C. The samples were held for about 20 minutes at this temperature for oxygenation and were then cooled quickly to room temperature (see Fig. 2). To investigate the potential of post-annealing steps to enhance grain boundary diffusion, several of the samples were post-annealed in 0.4 bar of O_2 at about 420 °C for half an hour. These additional anneals changed the sample properties only marginally. By wet etching, superconducting bridges with widths of 5–8 μm and lengths of 20 μm were patterned across the grain boundaries and inside the grains. Measurements at 77 K were done in a liquid nitrogen cryostat. The critical current densities were obtained from the ratios of the critical currents and the cross-sectional areas of the multilayer bridges, all of which were measured by scanning force microscopy. The entire film thickness, not just the thickness of the undoped $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ layers, was used in calculating J_c .

of doping heterostructures, several multilayer configurations and undoped control samples were investigated, fabricated on symmetric 24° [001] tilt SrTiO_3 bicrystals. In total, about 200 intergrain bridges on 40 substrates were analysed, as well as numerous bridges inside the grains. Details about the investigated configurations and the sample fabrication are given in Figs 1 and 2. To rule out inadvertent Ca doping of the $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ layers, a specially designed deposition system was used for the fabrication of the bilayer samples. This system was composed of two identical deposition chambers, one of which was dedicated to the growth of the Ca-free layers only. To keep this chamber completely free from calcium, separate heaters were used for the two chambers. Therefore, after the first layer was deposited, the samples were cooled to room temperature at 0.4 bar of O_2 , the vacuum was broken and the samples mounted onto another heater. To allow *in situ* growth and to limit the number of sample transfers, the trilayers and the superlattices were completely grown in the vacuum-chamber used for the deposition of the Ca-doped samples. In this case, a small amount of residual calcium present in the chamber was expected to be incorporated into the growing films and diffuses partially to the grain boundaries.

As anticipated, all grains except the homogeneously doped ones had critical temperatures above 90 K and critical current densities of approximately $6 \times 10^6 \text{ A cm}^{-2}$ at 77 K. All grain boundaries showed Josephson-junction type $I(V)$ characteristics with a distinct voltage step at J_c . Therefore, the intergrain critical currents could be unambiguously determined from the current–voltage characteristics, using a voltage criterion of 1 μV . The characteristic behaviour of the grain boundaries is presented in Fig. 3. This figure displays the $J_c(T)$ -dependencies of four samples: a $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ film, a homogeneously doped $\text{Y}_{0.7}\text{Ca}_{0.3}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ film, a bilayer consisting of a 130-nm $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ layer and a 20-nm-thick $\text{Y}_{0.7}\text{Ca}_{0.3}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ layer on top as illustrated in Fig. 1b, and a 20-nm $\text{Y}_{0.7}\text{Ca}_{0.3}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ /130-nm $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ /20-nm $\text{Y}_{0.7}\text{Ca}_{0.3}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ trilayer which had the highest J_c value of all the samples. As expected, the homogeneously doped $\text{Y}_{0.7}\text{Ca}_{0.3}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ sample shows a very large critical current density of approximately $7 \times 10^6 \text{ A cm}^{-2}$ at 4.2 K, but, as the Ca doping depresses its T_c , the sample is not superconducting at all at 77 K. At this temperature, however, the critical current densities of the 24° grain boundaries in the average doping bilayer and in the trilayer have already reached $1.2 \times 10^5 \text{ A cm}^{-2}$ and $3.3 \times 10^5 \text{ A cm}^{-2}$, greatly exceeding the J_c value of the undoped sample of $5.5 \times 10^4 \text{ A cm}^{-2}$. As the T_c of the individual layers doped with 30% Ca cannot exceed 77 K significantly, the supercurrent is presumably flowing completely in the $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ layer, which therefore in the trilayer sample has a net

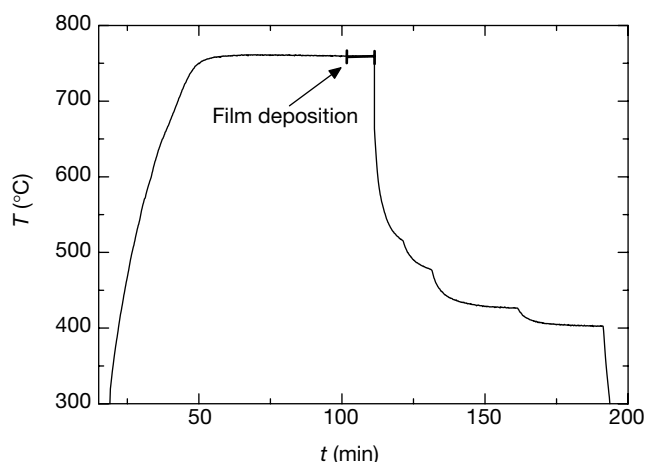


Figure 2 Measured temperature sequence of the substrate heater block used for the growth of the doping trilayers and superlattices.

critical current density of approximately $4.3 \times 10^5 \text{ A cm}^{-2}$ (77 K). These current densities equal those usually reported for 24° grain boundaries in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ films at 4.2 K (ref. 7) and are an order of magnitude larger than other 77 K data^{3,4}. Clearly, by combining the Ca-doped and the Ca-free films, both characterized by a small J_c value at 77 K, a multilayer system with an unprecedentedly high J_c has been obtained. The enhancement of the grain boundary J_c in comparison with the homogeneously doped samples occurs above a crossover temperature T^* , which is dependent on the sample design. For the bilayer and trilayer samples shown in Fig. 3, the crossover temperatures are 57 K and 52 K, respectively. We note that for any temperature a doping heterostructure exists which yields J_c values surpassing those of pure $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$.

To investigate the influence of the multilayer configuration on the J_c enhancement, several variations of the heterostructures were explored. First, bilayers with a total thickness t of 150–160 nm were fabricated and their J_c values measured as a function of the thickness d of the Ca-doped top layer. Figure 4 shows the critical current densities of the 85 superconducting bridges, with the error bars displaying the σ -spread of all samples fabricated. The values of d are given with an accuracy of 10%. The critical current density displays a maximum as a function of the top layer thickness and is enhanced by a factor of two for $d \approx 20$ nm. The J_c enhancement is smaller for thicker top layers, presumably because of the excessive doping. In addition to the overdoped samples, the J_c values of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}/\text{YBa}_2\text{Cu}_{2.9}\text{Co}_{0.1}\text{O}_{7-\delta}$ bilayers were also plotted in Fig. 4. Substitution of Cu by Co causes underdoping, and, as we expected, we measured a reduction of the critical current density.

Bilayers for which both t and d were scaled up by a factor of two display almost the same J_c increases achieved by standard bilayers. Likewise, after performing a post-anneal step, the bilayer J_c increased by only 3% at best. If the bilayer sequence was reversed, so that the undoped film was grown on top of the Ca-doped one, much smaller J_c enhancements were observed. Large critical current densities, however, have been obtained in superlattices. Superlattices, consisting of five and a half pairs of bilayers as sketched in

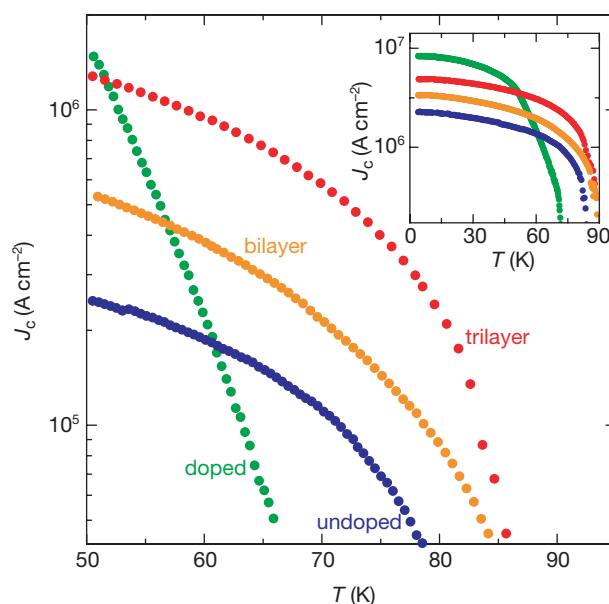


Figure 3 Temperature dependence of the critical current densities J_c of grain boundaries with various doping configurations, indicating the strong enhancement of J_c in doping heterostructures. Displayed are the $J_c(T)$ dependencies of symmetric 24° [001] tilt grain boundaries in various bicrystalline samples: a $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ film (undoped), a $\text{Y}_{0.7}\text{Ca}_{0.3}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ film (doped), a $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}/\text{Y}_{0.7}\text{Ca}_{0.3}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ bilayer with a J_c value typical of such bilayers, and a $\text{Y}_{0.7}\text{Ca}_{0.3}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}/\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}/\text{Y}_{0.7}\text{Ca}_{0.3}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ trilayer with an excellent critical current density.

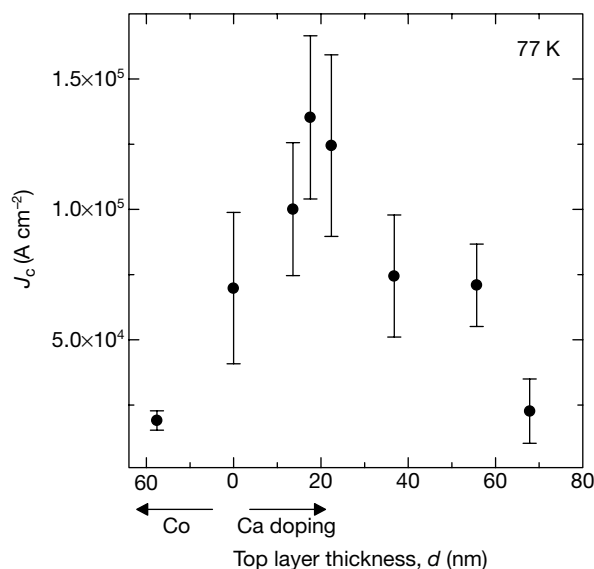


Figure 4 Dependence of the grain boundary critical current density J_c on the thickness of the doped top layer in doping bilayers. The figure shows the critical current density of 24° symmetric [001] tilt grain boundaries in doping bilayers about 160 nm thick at 77 K, plotted as a function of the thickness d of the doping top layer. On the right side of the figure, the data from $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}/\text{Y}_{0.7}\text{Ca}_{0.3}\text{Ba}_2\text{Cu}_3\text{O}_{7-x}$ bilayers are plotted. On the left side the data from $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}/\text{YBa}_2\text{Cu}_{2.9}\text{Co}_{0.1}\text{O}_{7-x}$ bilayers are shown. For the Co-doped samples the scale of the x-axis is compressed by a factor of three.

Fig. 1a, achieve critical current densities of $1.6 \times 10^5 \text{ A cm}^{-2}$ at 77 K for $x = 0.1$, $d = 25 \text{ nm}$, and a periodicity of 50 nm.

For several samples, the dependence of the critical current on the applied magnetic field H , oriented parallel to the c-direction of the films, was measured. The maximum fields applied were 250 μT . All samples showed the $I_c(H)$ dependencies expected for 24° grain-boundary Josephson junctions.

For the large-scale fabrication of coated-conductor-based superconducting cables, techniques such as ion beam assisted deposition (IBAD)^{11–13} or the rolling assisted biaxially aligned substrate process (RABITS)¹⁴ are being developed in order to enhance J_c by providing an optimal alignment of the grains^{2,3}. It is extremely desirable to lessen the required accuracy of the grain alignment. The use of doping heterostructures, which can easily be incorporated into these processes, may achieve this goal, as it presents a way of extending the established grain-boundary-angle-dependent limits on the critical current densities at 77 K (refs 2, 3). We anticipate that this will indeed be the case, considering that interface-charging and band-bending play an important role for small-angle grain boundaries as well^{8,9}. □

Received 21 March; accepted 16 August 2000.

1. Bednorz, J. G. & Müller, K. A. Possible high T_c superconductivity in the Ba-La-Cu-O system. *Z. Phys. B* **64**, 189–193 (1986).
2. Dimos, D., Chaudhari, P. & Mannhart, J. Superconducting transport properties of grain boundaries in $\text{YBa}_2\text{Cu}_3\text{O}_7$ bicrystals. *Phys. Rev. B* **41**, 4038–4049 (1990).
3. Ivanov, Z. G. *et al.* Weak links and dc SQUIDs on artificial nonsymmetric grain boundaries in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. *Appl. Phys. Lett.* **59**, 3030–3032 (1991).
4. Heinig, N. F., Redwing, R. D., Nordman, J. E. & Larbalestier, D. C. Strong to weak coupling transition in low misorientation angle thin film $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ bicrystals. *Phys. Rev. B* **60**, 1409–1417 (1999).
5. Mannhart, J. & Hilgenkamp, H. Wavefunction symmetry and its influence on superconducting devices. *Supercond. Sci. Technol.* **10**, 880–883 (1997).
6. Mannhart, J. & Hilgenkamp, H. Possible influence of band bending on the normal state properties of grain boundaries in high- T_c superconductors. *Mater. Sci. Eng. B* **56**, 77–85 (1998).
7. Hilgenkamp, H. & Mannhart, J. Superconducting and normal-state properties of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ bicrystal grain boundary junctions in thin films. *Appl. Phys. Lett.* **73**, 265–267 (1998).
8. Gurevich, A. & Pashitskii, E. A. Current transport through low-angle grain boundaries in high-temperature superconductors. *Phys. Rev. B* **57**, 13878–13893 (1998).
9. Schmehl, A. *et al.* Doping-induced enhancement of the critical currents of grain boundaries in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. *Europhys. Lett.* **47**, 110–115 (1999).

10. Ivanov, Z. G., Stepanov, E. A., Tzalenchuk, A. Y. & Claeson, T. Properties of locally doped bi-crystal grain boundary junctions. *Physica B* **194–196**, 2187–2188 (1994).
11. Iijima, Y., Tanabe, N., Kohno, O. & Ikeno, Y. In-plane aligned $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ thin films deposited on polycrystalline metallic substrates. *Appl. Phys. Lett.* **60**, 769–771 (1992).
12. Reade, R. P., Berdahl, P., Russo, R. E. & Garrison, S. M. Laser deposition of biaxially textured yttria-stabilized zirconia buffer layers on polycrystalline metallic alloys for high critical current Y-Ba-Cu-O thin films. *Appl. Phys. Lett.* **61**, 2231–2233 (1992).
13. Wu, X. D. *et al.* Properties of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ thick films on flexible buffered metallic substrates. *Appl. Phys. Lett.* **67**, 2397–2399 (1995).
14. Norton, D. P. *et al.* Epitaxial $\text{YBa}_2\text{Cu}_3\text{O}_7$ on biaxially textured nickel (001): an approach to superconducting tapes with high critical current density. *Science* **274**, 755–757 (1996).

Acknowledgements

We are grateful for valuable discussions with D. G. Schlom. This work was supported by the Bundesministerium für Bildung und Forschung. H.H. thanks the Royal Dutch Academy of Sciences for its support.

Correspondence and requests for materials should be addressed to J.M. (e-mail: jochen.mannhart@physik.uni-augsburg.de).

Thin films of fullerene-like MoS_2 nanoparticles with ultra-low friction and wear

Manish Chhowalla & Gehan A. J. Amaratunga

Engineering Department, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK

The tribological properties of solid lubricants such as graphite and the metal dichalcogenides MX_2 (where M is molybdenum or tungsten and X is sulphur or selenium)^{1–13} are of technological interest for reducing wear in circumstances where liquid lubricants are impractical, such as in space technology, ultra-high vacuum or automotive transport. These materials are characterized by weak interatomic interactions (van der Waals forces) between their layered structures, allowing easy, low-strength shearing^{14,15}. Although these materials exhibit excellent friction and wear resistance and extended lifetime in vacuum, their tribological properties remain poor in the presence of humidity or oxygen^{16–19}, thereby limiting their technological applications in the Earth's atmosphere. But using MX_2 in the form of isolated inorganic fullerene-like hollow nanoparticles similar to carbon fullerenes and nanotubes can improve its performance¹. Here we show that thin films of hollow MoS_2 nanoparticles, deposited by a localized high-pressure arc discharge method, exhibit ultra-low friction (an order of magnitude lower than for sputtered MoS_2 thin films) and wear in nitrogen and 45% humidity. We attribute this 'dry' behaviour in humid environments to the presence of curved S–Mo–S planes that prevent oxidation and preserve the layered structure.

Thus far, only isolated MoS_2 and WS_2 nanoparticles have been generated by solid–gas reaction or electron-beam irradiation^{20,21}. Here we report the friction and wear properties of thin films of MoS_2 nanoparticles, deposited using a localized high-pressure arc discharge^{22,23} (see Methods section). Films deposited by arc discharge, but in the absence of localized high pressure, appear mostly amorphous with only local hexagonal structure when studied by high-resolution electron microscopy (HREM). However, circular nanoparticles (Fig. 1a) and curved S–Mo–S planes (Fig. 1b) are readily seen for films deposited by arc discharge in the presence of high-pressure nitrogen. The key mechanism in the formation of nanoparticles is the bending and rearrangement of the basal planes²⁴. Displacement of Mo or S via collisions with energetic

ions in the arc plasma²⁵ can lead to rearrangement of atoms in order to maintain electrical neutrality. The results shown in Fig. 1 indicate that it is possible to generate well formed hollow fullerene-like 'onion' MoS₂ clusters in the form of a thin film using the localized arc discharge method.

In addition to HREM, further information regarding the chemical composition and structure was obtained from X-ray diffraction (XRD) analysis. The XRD pattern of the nanoparticle MoS₂ film is shown in Fig. 2. A sharp (0002) peak is observed, along with peaks from other reflection groups. The (0002) peak in the XRD spectrum of the nanoparticle film is characterized by a shift to a lower angle as

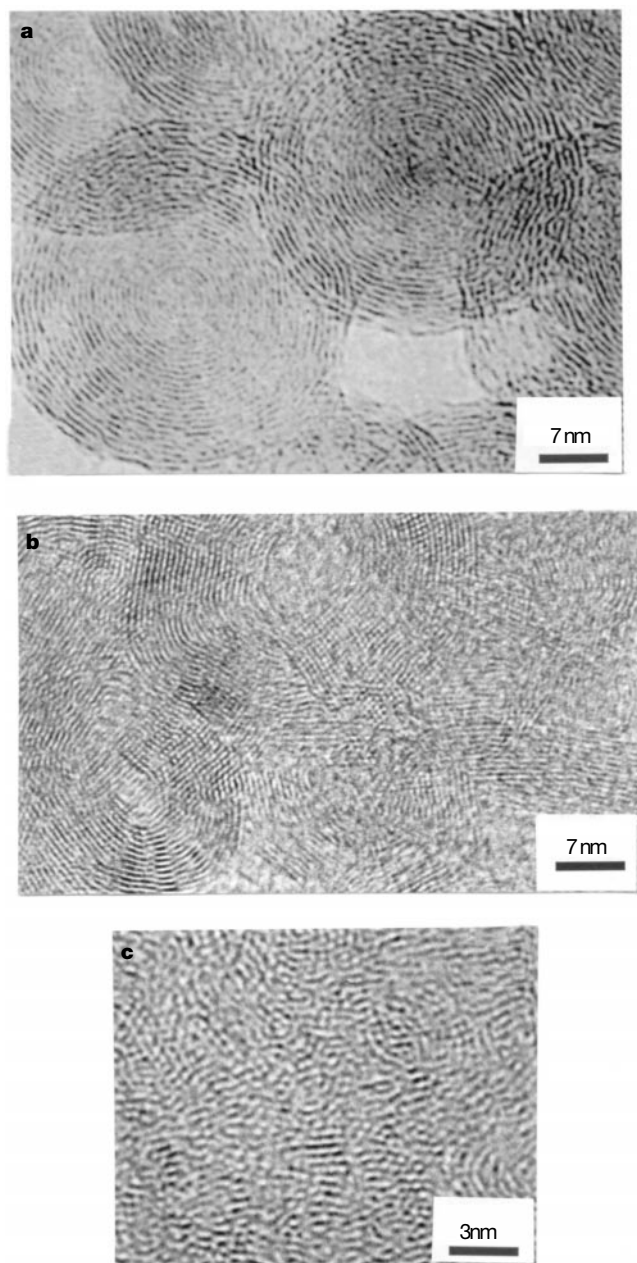


Figure 1 High-resolution transmission electron microscopy (TEM) images of 50-nm-thick films of MoS₂ nanoparticles. The films were deposited on Si[100] substrates, and the images were obtained using a JEOL 2000FX TEM operated at 200 kV. The TEM samples were prepared by dissolving the Si substrate in HF:HNO₃:H₂O solution (or in water for NaCl substrates), and placing the freed fragments of the film onto Cu grids. Circular nanoparticles with diameters up to 30 nm are readily visible in **a**. Curved S-Mo-S planes forming irregularly shaped nanoparticles are shown in **b**. An image of a sputtered MoS₂ film is also shown in **c**, for comparison.

compared to the (0002) peak in hexagonal (2H) MoS₂ crystals. This shift in the (0002) peak indicates lattice expansion, and has been attributed to the introduction of strain owing to curvature of the layers²⁴. In contrast, the (0002) peak shift in dense, oriented ion beam assisted deposition (IBAD). MoS₂ films has been attributed to straining of the basal plane spacing due to defects²⁶. In lamellar lattices, the (000 l) reflections are associated with ordering along the c -axis while the (hk 0) reflections indicate ordering in the basal planes. The prominence of the (0002) peak in the XRD pattern of the nanoparticle MoS₂ film in Fig. 2 indicates the presence of well-stacked layered structure. An estimation of the grain size from the full-width at half-maximum of the (0002) reflection yields about 30 nm, in agreement with the average diameter of the nanoparticles observed in HREM. In Fig. 2, we have also plotted for comparison the XRD spectra from MoS₂ films produced by vacuum arc and d.c. sputtering techniques.

The friction coefficient (μ) and wear rate (w) of the films of MoS₂ nanoparticles were determined by the ball-on-disk (pure sliding) method^{11,27}. The disks were the same ones used to measure the hardness and adhesion (see Methods). The results from the different types of films tested in our laboratory are summarized in Table 1. The evolution of μ as a function of the number of rubbings is plotted in Fig. 3. The nanoparticle MoS₂ films show a very low friction coefficient in a dry nitrogen atmosphere ($\mu = 0.006$). But our central result is that μ and w remain ultra-low even in 45% relative humidity in ambient conditions. In fact, μ and w are slightly lower for the nanoparticle MoS₂ film in 45% humidity than for the sputtered MoS₂ film in dry nitrogen.

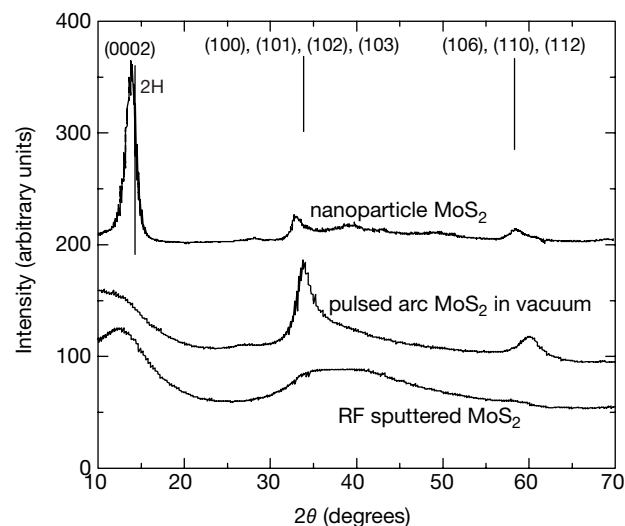


Figure 2 X-ray diffraction (XRD) patterns from thin films of MoS₂ nanoparticles. The figure shows results for three films: one produced by a localized high-pressure discharge, one deposited in vacuum, and one deposited by sputtering. The XRD was performed, using a Siemens MRD diffractometer with a Cu source, on 1.2- μ m-thick films deposited on 440C stainless steel disks. In order to enhance the signal from the thin films, all measurements were performed at a grazing angle of $\alpha = 3^\circ$. XRD spectra obtained from films deposited on background-free quartz also revealed a similar pattern. The nanoparticle MoS₂ film shows a sharp (0002) peak, in contrast to the vacuum-arc and sputter-deposited films in which other reflections dominate the XRD pattern. The slight shift of the (0002) peak indicated by the vertical line is explained in the text.

Table 1 Results of tribological tests on MoS₂ films

Film	μ	w (mm ³ N ⁻¹ mm ⁻¹)
Nanoparticle MoS ₂	0.008–0.01	$\sim 1 \times 10^{-11}$
Sputtered MoS ₂	0.1–0.3	$\sim 3 \times 10^{-9}$
Hard TiN	0.4–0.6	NA

μ is measured friction coefficient in 45% relative humidity in ambient conditions; w is the wear rate. NA, not available.

Examinations of the wear patterns (Fig. 3b,c) on the uncoated ball show clear differences between the nanoparticle and the sputtered films. In the sputtered film, a heavily worn and rough surface with bright and shiny wear tracks along with some dark patches can be seen (Fig. 3c). The wear pattern on the uncoated ball after failure was also examined under a scanning electron microscope fitted with energy dispersive X-ray (EDX) detector. EDX analysis revealed predominantly iron oxide in the wear scar. X-ray photoelectron spectroscopy (XPS) analysis of the wear region on the coated disk also confirmed the formation of MoO_3 on the sputtered film (after failure). The test on the nanoparticle film was stopped before failure after 2.4×10^5 rubbings (three times longer than the rubbings to failure in the sputtered film) owing to the limitations of our experimental apparatus. In contrast to Fig. 3c, substantial wear in the form of bright and dark patches was not observed in the scar of the ball mating with the nanoparticle MoS_2 film (Fig. 3b). Instead, substantial film transfer from the coated disk to the uncoated ball can be seen. EDX analysis of the worn region in Fig. 3b showed that primarily molybdenum and sulphur were present, with only residual peaks arising from carbon and oxygen. The XPS analysis of the worn region on the nanoparticle MoS_2 film before failure revealed a smaller MoO_3 peak than the sputtered film.

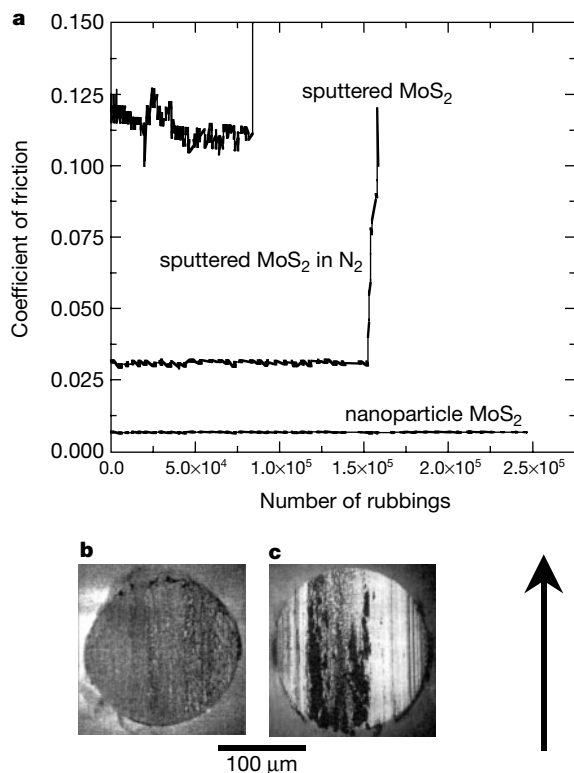


Figure 3 Wear tests on MoS_2 films in different atmospheres. **a**, The coefficient of friction as a function of time for: sputtered MoS_2 films in 45% relative humidity in ambient conditions (top curve); sputtered MoS_2 films in dry nitrogen; and nanoparticle MoS_2 films in 45% humidity formed by the procedure described in this work. Note that each curve is an average of ten different measurements on at least three different samples. The friction coefficient was obtained by a unidirectional 'ball on flat disk' (pure sliding) wear test (see Methods). The mating ball was also made from 440C stainless steel with a diameter of 0.7 cm and of comparable average roughness to the disk. The applied load was 10 N (using a dead-weight) on to the rotating disk (speed, 50 cm s^{-1}) which gave a maximum hertzian pressure of $\sim 1.1 \text{ GPa}$ and contact diameter of $\sim 100 \mu\text{m}$. Calibration of the ball-on-disk was performed to ensure that the tilt of the disk was within 0.5° , which gave a tolerance of ± 0.009 for the friction coefficient. The mating ball was uncoated. The wear rate for the disk was obtained by determining the cross-sectional area of the wear track. The wear patterns for nanoparticle (**b**) and sputtered (**c**) MoS_2 films were obtained using a light optical microscope. The arrow indicates the direction of sliding.

In order to study the structural changes in the film after the test, MoS_2 debris from the wear scar of the nanoparticle film was observed under HREM. The observations revealed that although the curved hexagonal lattice is still maintained (Fig. 4), almost no closed circular nanoparticles remained, indicating that under these loading conditions the large fullerenes fragment into smaller irregularly shaped closed crystallites. An HREM image of the nanoparticle film after 2.4×10^5 rubbings is shown in Fig. 4.

Although the tribological properties of MoS_2 powder and MoS_2 thin films have been investigated for decades, the mechanisms responsible for the lubrication process remain unclear^{4,6–9,11}. The combination of factors such as maintenance of the lamellar structure, formation and adhesion of a homogeneous transfer film, prevention of oxidation and intercrystallite slip have been cited as possible mechanisms responsible for the successful performance of MoS_2 as a solid boundary lubricant^{4–11}. In the case of WS_2 nanoparticles dispersed in oil, rolling friction along with the absence of dangling bonds may be responsible for the low μ and w (ref. 1). In contrast, we believe that the mechanisms responsible for the low friction coefficient in fullerene-like MoS_2 films are probably similar to sputtered films but with additional features. The tribological properties of sputtered MoS_2 films are correlated to its microstructure, and the variations are attributed to differences in deposition parameters or the degree of water vapour present during processing¹¹. Annealing of sputtered MoS_2 films can lead to the sharpening of the (0002) XRD peak, as well as the appearance of short, two-dimensional hexagonal crystals in HREM which in turn has been correlated to the decrease in the friction coefficient²⁸. Therefore, it is feasible that highly ordered and pure MoS_2 could have an intrinsically low friction coefficient if the low-shear-strength properties are maintained during tribological testing. In our case, the coating process was initiated at a base pressure of $1 \times 10^{-5} \text{ Pa}$ and the substrates were held at 200°C , so contamination during deposition is insignificant. The HREM images (Fig. 1a, b) primarily show circular nanoparticles and curved S–Mo–S planes, along with fine amorphous regions, before testing. Furthermore, the XRD shows strong ordering in the c -axis.

Taking into account the HREM, XRD, EDX and XPS data, our results clearly indicate that it is possible to improve the friction and wear properties of MoS_2 thin films by the incorporation of fullerene-like nanoparticles. We argue that the low friction coefficient measured for the nanoparticle films may simply be a consequence of the highly ordered curved structure. That is, the incorporation of nanoparticles through localized high-pressure arc discharge allows the deposition of a highly ordered MoS_2 structure that is not (or may not be) accessible by sputtering or other methods²⁹. Further-

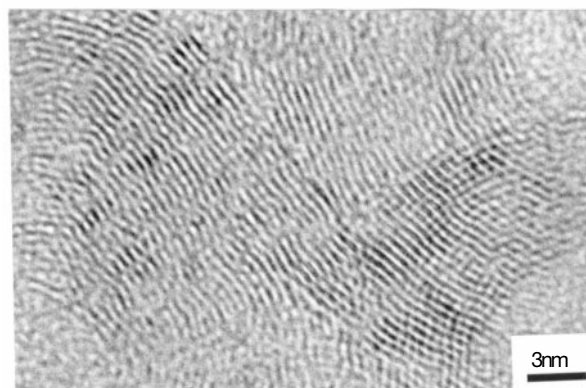


Figure 4 HREM image of debris collected from the wear track region of an MoS_2 nanoparticle film after the reciprocating wear test in 45% relative humidity in ambient conditions. The image clearly shows the presence of curved hexagonal planes.

more, mechanisms such as intercrystallite slip can still occur in fragmented nanoparticles and crystallites consisting of irregularly shaped curved S–Mo–S planes, even though shearing between the layers is unlikely in closed nanoparticles. An additional feature of these films is the presence of curved hexagonal planes that clearly enhance the stability of the material in humid environments, and thus help to preserve the lamellar structure longer. In the case of MoS₂ powder (or sputtered film), oxidation can occur through uninterminated bonds of the hexagonal planes, causing the material to stick, leading to rapid deterioration of its low-shear-strength properties. The curvature of the hexagonal planes appears to buffer the Mo atom from oxidation by reducing the number of exposed dangling bonds at the edges of the planes.

The deposition of films of MoS₂ nanoparticles could be easily incorporated into existing thin-film technology for coating automotive parts and tool components. The deposition method that we report here allows the formation of the low-friction layer directly onto the components instead of dispersing the nanoparticles into oil, which could lead to adverse clogging. The low-temperature nature of the process also allows it to be used on top of hard, or magnetic, thin films to give a low-friction layer. Finally, we note that the ultra-low friction coefficients measured here may point towards the intrinsic tribological properties of highly ordered solid MoS₂, but additional fundamental study is required to test this hypothesis. □

Methods

Deposition of nanoparticle MoS₂

The nanoparticle MoS₂ films were deposited by ablating a solid MoS₂ target by an arc discharge in the presence of localized high-pressure nitrogen, similar to the technique used to generate thin films of nanoparticle carbon^{22,23}. The localized high pressure is generated by introducing the carrier gas via a 1-mm hole in the MoS₂ target. The dynamics involved in the formation of MoS₂ nanoparticles are at present unclear but the mechanism is probably similar to the formation of carbon nanoparticles, as the technique used here is identical and is described in detail elsewhere^{22,23}. The nanoparticles of MoS₂ are thought to form immediately above the cathode surface, and to be carried via expansion from the localized high-pressure region near the arc discharge to the substrate; the substrate is placed 20 cm away, where the pressure is kept constant at 10 mtorr (refs 22,23). The films were deposited at 200 °C. The arc was ignited (at 75 A and ~22 V) in the localized region of high-pressure nitrogen.

Mechanical properties of nanoparticle MoS₂

The films were deposited on disks of 440C stainless steel; these disks were 5 cm in diameter, 0.8 cm thick, and were used for all the mechanical and tribological measurements reported here. The substrates were lap-polished so that the final average roughness value was 0.04–0.05 µm. The film thickness was kept constant at 1.2 ± 0.1 µm for all the mechanical and tribological tests. The adhesion, measured using a VTT scratch tester with a 200-µm Rockwell C indenter and maximum load of 100 N with a load rate of 1.7 N s⁻¹, was found to be 25 N. The hardness of the films was measured using a Fisher microindentation system with Berkovich (pyramidal) tip²². The hardness, extracted using the plastic displacement obtained from the intercept with the displacement axis of the unloading curve at maximum applied load (5 mN), was found to be 10 GPa, comparable to values reported for sputtered MoS₂ (ref. 30). The surface roughness of the deposited film measured using the Dektak IIA profilometer was found to be approximately 30 nm.

Received 25 October 1999; accepted 20 June 2000.

1. Rapoport, L. *et al.* Hollow nanoparticles of WS₂ as potential solid-state lubricants. *Nature* **387**, 791–793 (1997).
2. Bell, M. E. & Findlay, J. H. Molybdenite as a new lubricant. *Phys. Rev.* **59**, 922–927 (1941).
3. Braithwaite, E. R. & Rowe, G. W. Principles and applications of lubrication with solids. *Sci. Lubricat.* **15**, 92–96 (1957).
4. Winer, W. O. Molybdenum disulphide as a lubricant: A review of the fundamental knowledge. *Wear* **10**, 422–452 (1967).
5. Spalvins, T. Lubrication with sputtered MoS₂ films. *ASLE Trans.* **14**, 267–271 (1972); Morphological and frictional behavior of sputtered MoS₂ films. *Thin Solid Films* **96**, 17–24 (1982).
6. Holinski, R. & Gansheimer, J. A study of the lubricating mechanism of molybdenum disulphide. *Wear* **19**, 329–342 (1972).
7. Buck, V. Morphological properties of sputtered MoS₂ films. *Wear* **91**, 281–288 (1983).
8. Fleischauer, P. D. & Tolentino, L. U. Effects of crystallite orientation on the environmental stability and lubrication properties of sputtered MoS₂ thin films. *ASLE Trans.* **27**, 82–88 (1984).
9. Roberts, E. W. Thin solid lubricants in space. *Tribol. Int.* **23**, 95–104 (1990).
10. Miyoshi, K. *et al.* A vacuum (10⁻⁹ Torr) friction apparatus for determining friction and endurance like that of MoS₂ films. *Tribol. Trans.* **36**, 351–358 (1993).
11. Suzuki, M. Comparison of tribological characteristics of sputtered MoS₂ films coated with different apparatus. *Wear* **218**, 110–118 (1998).
12. Wang, D. Y., Chang, C. L., Chen, Z. Y. & Ho, W. Y. Microstructural characterization of MoS₂-Ti composite solid lubricating films. *Surf. Coat. Technol.* **121**, 629–635 (1999).

13. Killeffer, D. H. & Linz, D. *Molybdenum Compounds: Their Chemistry and Technology* (Interscience, New York, 1952).
14. Dickinson, R. G. & Pauling, L. The crystal structure of molybdenite. *J. Am. Chem. Soc.* **45**, 1466–1471 (1923).
15. Bragg, W. The investigation of the properties of thin films by means of x-rays. *Nature* **115**, 266–269 (1925).
16. Williams, J. A. *Engineering Tribology* 368–370 (Oxford Univ. Press, 1994).
17. Savage, R. H. Graphite lubrication. *J. Appl. Phys.* **19**, 1–10 (1948).
18. Arnell, R. D., Davies, P. B., Halling, J. & Whomes, T. L. *Tribology—Principles and Design Applications* 105–107 (Macmillan, London, 1991).
19. Matsumoto, K. & Suzuki, M. in *Proc. Int. Tribology Conf.* 1995 1165–1169 (JST, Tokyo, 1996).
20. Feldman, Y., Wasserman, E., Srolovitz, D. J. & Tenne, R. High-rate, gas-phase growth of MoS₂ nested inorganic fullerenes and nanotubes. *Science* **267**, 222–225 (1995).
21. Jose-Yacamán, M. *et al.* Studies of MoS₂ structures produced by electron irradiation. *Appl. Phys. Lett.* **69**, 1065–1067 (1996).
22. Amarutunga, G. A. *et al.* Hard elastic carbon thin films from linking of carbon nanoparticles. *Nature* **383**, 321–323 (1996).
23. Chhowalla, M., Aharonov, R. A., Kiely, C. J., Alexandrou, I. A. & Amarutunga, G. A. J. Generation and deposition of fullerene- and nanotube-rich carbon thin films. *Phil. Mag. Lett.* **75**, 329–335 (1997).
24. Srolovitz, D. J., Safran, S. A., Homyonfer, M. & Tenne, R. Morphology of nested fullerenes. *Phys. Rev. Lett.* **74**, 1779–1782 (1995).
25. Bilek, M. M. M., Martin, P. J. & McKenzie, D. R. Influence of gas pressure and cathode composition on ion energy distributions in filtered vacuum arcs. *J. Appl. Phys.* **83**, 2965–2970 (1998).
26. Dunn, D. N., Seitzman, L. E. & Singer, I. L. The origin of anomalous low (0002) peak in x-ray diffraction spectra of MoS₂ films grown by ion beam assisted deposition (IBAD). *J. Mater. Res.* **12**, 1191–1194 (1997).
27. Westergaard, R. & Axén, N. *Tribologia—Finn. J. Tribol.* **17**, 14–18 (1998).
28. Mattern, N., Hermann, H., Weise, G., Teresiak, A. & Bauer, H. D. Structure and properties of MoS₂ films. *Mater. Sci. Forum* **235–238**, 613–618 (1997).
29. Dunn, D. N., Seitzman, L. E. & Singer, I. L. MoS₂ deposited by ion beam assisted deposition: 2H or random layer structure? *J. Mater. Res.* **13**, 3001–3007 (1998).
30. Fox, V. C., Renevier, N., Teer, D. G., Hampshire, J. & Rigato, V. The structure of tribologically improved MoS₂-metal composite coatings and their industrial applications. *Surf. Coat. Technol.* **119**, 492–497 (1999).

Acknowledgements

We thank R. Aharonov for general assistance and advice, and J. A. Williams for discussions.

Correspondence and requests for materials should be addressed to G.A.J.A. (e-mail: ga@eng.cam.ac.uk).

Helical self-assembled polymers from cooperative stacking of hydrogen-bonded pairs

J. H. K. Ky Hirschberg, Luc Brunsveld, Aissa Ramzi, Jef A. J. M. Vekemans, Rint P. Sijbesma & E. W. Meijer

Laboratory of Macromolecular and Organic Chemistry and Dutch Polymer Institute, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands

The double helix of DNA epitomizes this molecule's ability to self-assemble in aqueous solutions into a complex chiral structure using hydrogen bonding and hydrophobic interactions. Non-covalently interacting molecules in organic solvents are used to design systems that similarly form controlled architectures^{1–7}. Peripheral chiral centres in assemblies^{8,9} and chiral side chains attached to a polymer backbone^{10,11} have been shown to induce chirality at the supramolecular level, and highly ordered structures stable in water are also known^{12–15}. However, it remains difficult to rationally exploit non-covalent interactions for the formation of chiral assemblies that are stable in water, where solvent molecules can compete effectively for hydrogen bonds. Here we describe a general strategy for the design of functionalized monomer units and their association in either water or alkanes into non-covalently linked polymeric structures with controlled helicity and chain length. The monomers consist of bifunctionalized ureidotriazine units connected by a spacer and

carrying solubilizing chains at the periphery. This design allows for dimerization through self-complementary quadruple hydrogen bonding between the units and solvophobic induced stacking of the dimers into columnar polymeric architectures, whose structure and helicity can be adjusted by tuning the nature of the solubilizing side chains.

The experimental verification of our design strategy is based on structures 1–4 (Fig. 1), which offer unprecedented control over the hierarchical growth of supramolecular organization. The synthesis of the mono- and bifunctional molecules is based on the acylation of the corresponding diaminotriazines with butylisocyanate and diisocyanatohexane, respectively. All compounds persist as single molecules in dimethylsulphoxide, as shown by their ^1H NMR spectra. But in chloroform, strong cooperative hydrogen bonding is favoured, leading to the dimerization of the ureidotriazine units via quadruple hydrogen bonds with an association constant (K_{ass}) of $2 \times 10^4 \text{ M}^{-1}$ (ref. 16). Compounds 2 and 4 thus assemble in chloroform into supramolecular random coil polymers, resulting in viscous solutions (a $3.1 \times 10^{-2} \text{ M}$ solution of 2a has a relative viscosity (η_{rel}) of 3.94).

Additional to the hydrogen bonding, solvophobic interactions between the aromatic ‘surfaces’ occur in 1 and 2 in nonpolar solvents. Association via the hydrogen bonding described above leads to the formation of dimers with a large and planar aromatic core surrounded by six flexible alkyl chains, a structure that is conducive to the formation of aggregates with a columnar polymeric architecture in environments that are unable to solvate the core¹⁷. ^1H NMR spectra of 1 and 2 recorded in deuterated chloroform show the typical resonances for the hydrogen-bonded structure at chemical shifts (δ) of 10.3, 9.9 and 9.3 p.p.m., while in deuterated dodecane only broad unresolved signals for the aliphatic tails and very broad and hardly detectable signals for the aromatic core of the molecules are observed.

We used circular dichroism (CD) spectroscopy of the chiral bifunctional compound 2b to probe the degree of order induced by the ensemble of interactions. In dodecane solutions, two CD bands were observed (Fig. 2a inset), with maxima corresponding to the UV absorption maxima. The Cotton effect in the absorption bands of the aromatic core of 2b is due to a helical arrangement of chromophores with a preferred handedness. The $\Delta\epsilon$ values of the CD spectrum decreased by only a factor of 2 when the concentration was lowered from 10^{-3} to 10^{-6} M (Fig. 2a). Hence, down to 10^{-6} M , most of the molecules are part of helical columns. A plot of the Cotton effect against temperature (Fig. 2b) shows that all helicity is lost over a relatively narrow temperature range from 60 to 80 °C, indicating that thermal denaturation is a cooperative process. This process is fully reversible, and identical CD spectra were observed

after cooling without appreciable hysteresis. The helical architectures can also be denatured with solvent. Addition of 2.6% of chloroform to a dodecane solution of 2b decreased the $\Delta\epsilon$ by a factor of 10, while in pure chloroform no Cotton effect was observed. The loss of chirality is due to disruption of the stacking interactions between the chromophores in chloroform. That the presence of a covalent linker between the disks is important in maintaining supramolecular chirality, is shown by the absence of a Cotton effect in solutions of monofunctional 1b, not only in chloroform, but also in hexane. In the latter solvent the dimers of 1b are stacked, but the columns lack the hexamethylene chain that induces positional order between disks of 2b. Within the columns formed by 1b, the dimerized molecules are therefore rotating freely.

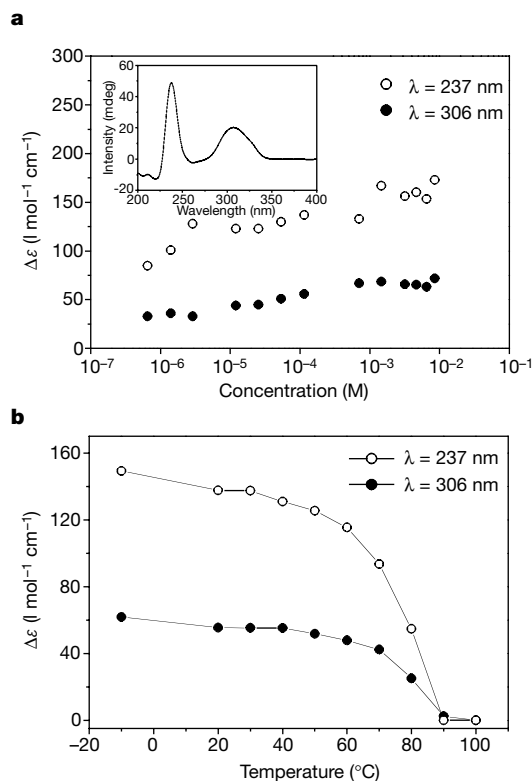


Figure 2 Circular dichroism spectra of 2b in dodecane. **a**, Concentration dependence of $\Delta\epsilon$ ($\Delta\epsilon$ is the Cotton effect normalized to concentration). Inset, typical CD spectrum. **b**, Temperature dependence of $\Delta\epsilon$ at two wavelengths at a concentration of $1.2 \times 10^{-4} \text{ M}$. Samples were kept for 10 min at the desired temperature before measurement.

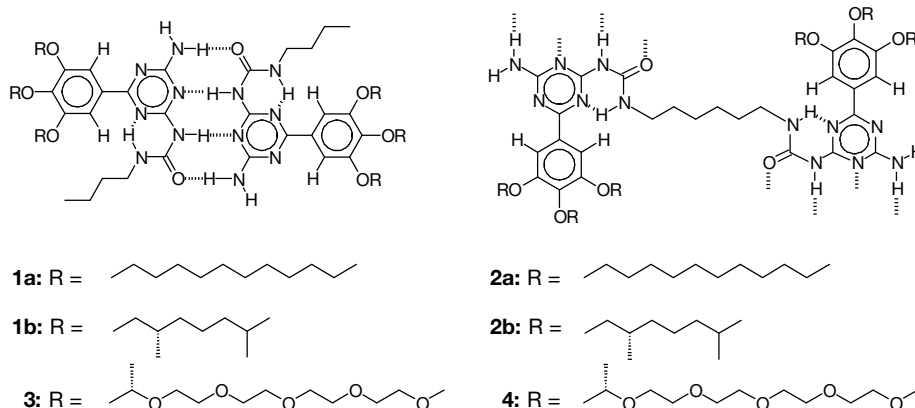


Figure 1 Structures of compounds 1–4. This figure shows the monofunctional ureidotriazines 1 and 3 and their mode of association via quadruple hydrogen bonds.

Structures of bifunctional derivatives 2 and 4 are also shown.

Small-angle neutron scattering (SANS) experiments prove the presence of the columnar architectures in deuterated dodecane solutions for both **1b** and **2a**. (In deuterated chloroform no large particle scattering is observed for **1b**, since the monofunctional molecules only exist as dimers.) We have fitted the scattering patterns obtained for **1b** with a model for cylindrical-like structures. The radius of the columns is 15 ± 1 Å and independent of concentration, consistent with a column built up of stacked dimers. The length of the columns is concentration dependent, and increases from 100 Å for a 0.2 wt% solution to 190 Å for a

1.0 wt% solution (about 60 molecules). At about 100 °C the scattered intensity decreases strongly, indicating the dissociation of the aggregates: the molecules then only exist as dimers, as in chloroform at room temperature. We have used this model to interpret the scattering patterns of **2a** in dodecane, although this is hampered by intercolumnar interactions (that is, gelation) at higher concentrations. The constant radius of the cylinders is now 17 ± 1 Å, larger than for **1b**, in agreement with the difference in side-chain length of **a** and **b**. In Fig. 3 we give the results of a chain-stopper experiment; monofunctional **1b** was mixed with

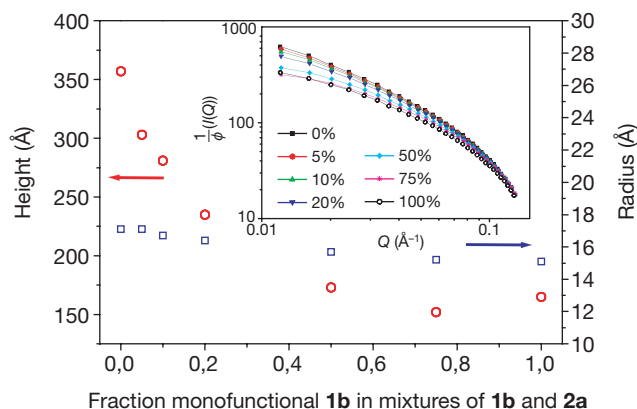


Figure 3 Small-angle neutron scattering experiments. Main figure, the height (circles) and radius (squares) of columns formed by **2a** as a function of the fraction of monofunctional **1b**. The concentration of **2a+1b** is kept constant at 0.5 wt% in the

dodecane solvent. Inset, the observed scattering patterns in a logarithmic plot of the normalized intensity (intensity $I(Q)$ divided by volume fraction ϕ) versus the scattering vector (Q), at different percentages of monofunctional **1b** in mixtures of **1b** and **2a**.

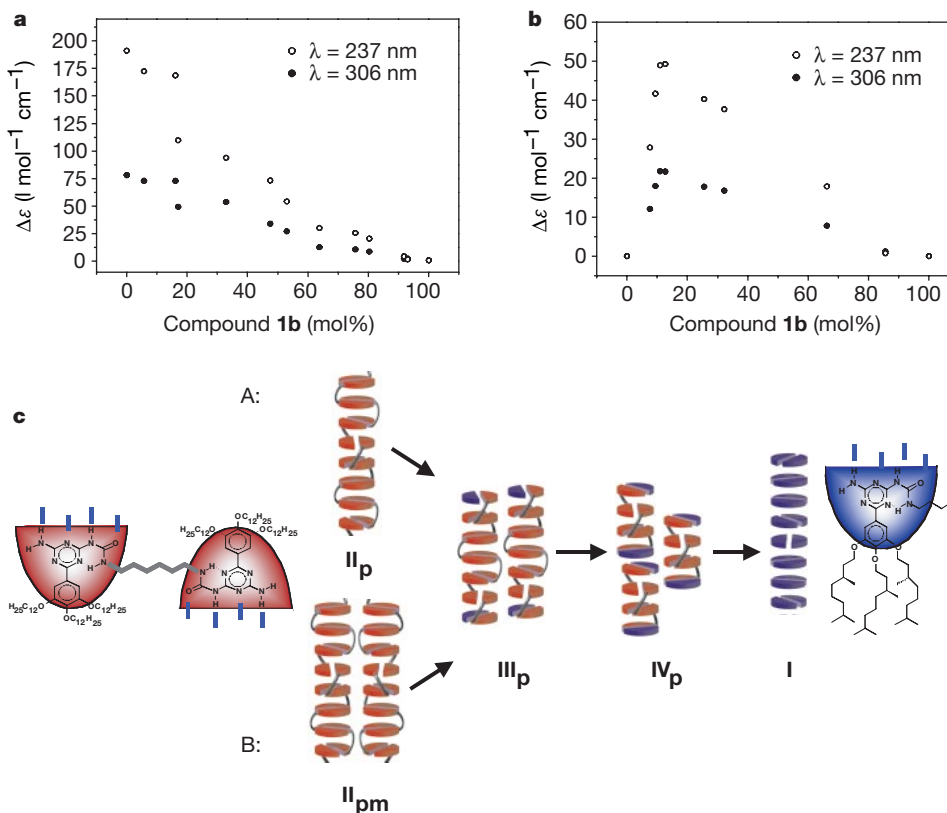


Figure 4 Intensity of the Cotton effect of mixtures of mono- and bifunctional compounds at two wavelengths. **a**, Chiral compounds **2b** (red) and **1b** (blue) at a constant total concentration of 0.02 wt%. **b**, Chiral monofunctional compound **1b** and achiral bifunctional compound **2a** at a constant total concentration of 0.02 wt%. **c**, A schematic representation of the organization of mixtures of mono- and bifunctional compounds upon increasing the fraction of monofunctional compound. In **I**, columns are formed by the

stacking of dimers. In **II** (with p and m corresponding to the right- and left-handed helical conformation, respectively), consecutive disks of dimeric ureidotriazines are connected by hexamethylene spacers. Transformations A and B correspond to the CD measurements in **a** and **b**, respectively. The helical sense of the biased columns (p) has been chosen arbitrarily: we cannot at present determine the handedness of the columns from the Cotton effect.

bifunctional **2a** in deuterated dodecane, while maintaining a constant 0.5 wt% concentration. The radius of the columns slowly changes from 17 to 15 Å as expected, but the length decreases rapidly after addition of small amounts of **1b**, showing the formation of shorter and less stable columns than pure bifunctional **2a**. The length of the columns is shorter for **1b** due to the absence of the supramolecular polymeric backbone.

The SANS results can be used to explain the observed Cotton effects of mixtures of **1b** and **2b** (Fig. 4a). Addition of end capper **1b** to a solution of **2b** results in the linear decrease of the Cotton effect. The addition of **1b** shortens the polymer chains. Thus, even though columns remain present, they consist of short polymer chains stacked on top of one another without long-range helical order, and this leads to a decrease in $\Delta\epsilon$. The columns formed by **2** feature a very high degree of cooperative order and their conformational state is fully controlled, as shown by adding chiral monofunctional compound **1b** to achiral **2a** (Fig. 4b). Both compounds separately are optically inactive in dodecane solution. However, in mixed solutions a strong Cotton effect is observed. Although they form just the ends of the polymer chains, the molecules of **1b** are still capable of transferring their chirality into the column and thus bias their helicity through cooperative interactions, resulting in a significant amplification of chirality. A cartoon representation of the stacks is given in Fig. 4c.

Solutions of **2a** in dodecane at higher concentrations are highly viscous and show pronounced shear thinning (see Supplementary Information), indicating the presence of lateral interactions (physical cross-links) between the columns. In contrast, a 4.8×10^{-3} M solution of **1a** has a viscosity similar to the solvent viscosity. At high concentrations of **1b** and **2a** in dodecane, the lateral interactions between the columns cause lyotropic mesophases to form. At room temperature, a solution of **1b** in dodecane becomes anisotropic at 0.25 M (20 wt%) while a solution of **2a** becomes anisotropic at 0.024 M (5 wt%). The columnar order of **1b** and **2a** is retained in the bulk as elucidated by optical microscopy, differential scanning calorimetry and X-ray diffraction. Chiral monofunctional derivative **1b** displays a mesophase with conical fan shaped texture (indicative of a columnar phase) between -25°C and 180°C . Compound **2a** exhibits a similar mesophase between -26°C and 178°C . The X-ray diffraction pattern of the mesophase of **2a** shows that the columns pack pseudohexagonally in a rectangular lattice with an intercolumnar distance of 35.2 Å.

The creation of similar columnar architectures in water requires the presence of a hydrophobic microenvironment that shields the hydrogen bonding from the solvent and makes them operative. We thought that the solvophobic interactions between the aromatic surfaces would produce such a hydrophobic microenvironment, and that by providing the bifunctional monomers with chiral ethylene oxide side chains, solubility in water would be ensured.

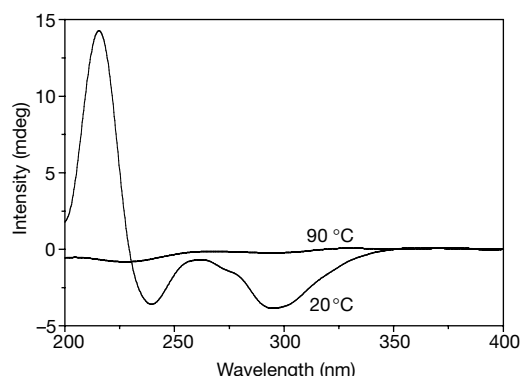


Figure 5 CD spectra of **4** in water, 2×10^{-4} M at 20°C and 90°C . The Cotton effect is independent of cooling or heating rate and reversible and stable in time.

Compounds **3** and **4** were, therefore, designed and synthesized in an enantiomerically pure form. Like its nonpolar analogue **2**, **4** is thermotropic liquid crystalline and forms a random coil supramolecular polymer in chloroform solutions. The polymer structure is destroyed in dimethylsulphoxide, since this solvent interferes with the hydrogen bonding. In water, however, **4** does form a helical column; this helicity is biased by the peripheral chiral moieties, as evidenced by a Cotton effect in the chromophores of the aromatic core at 10^{-4} M (Fig. 5). The CD spectrum of **4** features the same bandshape characteristics as those of **2b** in alkane solvents, indicating that a similar helical structure is formed. The Cotton effect of **4** in water is opposite to that of **2b** in alkanes, indicating opposite preferred helicities. At 90°C , no positional order is present in the architecture, as shown by the absence of a Cotton effect (Fig. 5). At higher temperatures, the water apparently overcomes the aggregation and interferes with the hydrogen bonds. Like its nonpolar analogue, chiral monofunctional water soluble **3** does not form helical structures in water, as evidenced by the absence of a Cotton effect at 10^{-4} M at room temperature. A supramolecular polymer backbone is necessary to form stable, well-defined helices: stacking of the hydrogen-bonded pairs leads to hydrophobic microdomains, allowing cooperative hydrogen bonding and transfer of the peripheral chirality into the helix.

The similar findings that we report here for **3** and **4** in both water and dodecane show the generality of our design criteria, and offer a unique approach to control the hierarchical growth of self-assembled structures. □

Received 14 March; accepted 2 August 2000.

- Whitesides, G. M., Mathias, J. P. & Seto, C. T. Molecular self-assembly and nanochemistry—A chemical strategy for the synthesis of nanostructures. *Science* **254**, 1312–1319 (1991).
- Krzywicki, F. G., Fouquey, C. & Lehn, J.-M. Electron-microscopic study of supramolecular liquid-crystalline polymers formed by molecular-recognition-directed self-assembly from complementary chiral components. *Proc. Natl Acad. Sci. USA* **90**, 163–167 (1993).
- Sijbesma, R. P. *et al.* Reversible polymers formed from self-complementary monomers using quadruple hydrogen-bonding. *Science* **278**, 1601–1604 (1997).
- Castellano, R. K., Rudkevich, D. M. & Rebek, J. Polycaps—Reversibly formed polymeric capsules. *Proc. Natl Acad. Sci. USA* **94**, 7132–7137 (1997).
- Gellman, S. H. Foldamers—a manifesto. *Acc. Chem. Res.* **31**, 172–180 (1998).
- Seebach, D. *et al.* Beta-peptides—synthesis by Arndt-Eistert homologation with concomitant peptide coupling—structure determination by NMR and CD spectroscopy and by X-ray crystallography—helical secondary structure of a beta-hexapeptide in solution and its stability towards pepsin. *Helv. Chim. Acta* **79**, 913–941 (1996).
- Nelson, J. C., Saven, J. G., Moore, J. S. & Wolynes, P. G. Solvophobic driven folding of nonbiological oligomers. *Science* **277**, 1793–1796 (1997).
- Palmans, A. R. A., Vekemans, J. A. J. M., Havinga, E. E. & Meijer, E. W. Sergeants-and-soldiers principle in chiral columnar stacks of disc-shaped molecules with C_3 symmetry. *Angew. Chem. Int. Edn Engl.* **36**, 2648–2651 (1997).
- Prins, L. J., Huskens, J., de Jong, F., Timmerman, P. & Reinhoudt, D. N. Complete asymmetric induction of supramolecular chirality in a hydrogen-bonded assembly. *Nature* **398**, 498–502 (1999).
- Green, M. M. *et al.* A helical polymer with a cooperative response to chiral information. *Science* **268**, 1860–1866 (1995).
- Prince, R. B., Brunsveld, L., Meijer, E. W. & Moore, J. S. Twist sense bias induced by chiral side chains in helically folded oligomers. *Angew. Chem. Int. Edn Engl.* **39**, 228–230 (2000).
- Gottarelli, G. *et al.* The self-recognition and self-assembly of folic-acid salts in isotropic water solution. *Helv. Chim. Acta* **79**, 220–234 (1996).
- Appella, D. H., Barchi, J. J. Jr, Durell, S. R. & Gellman, S. H. Formation of short, stable helices in aqueous solution by β -amino acid hexamers. *J. Am. Chem. Soc.* **121**, 2309–2310 (1999).
- Blokzijl, W. & Engberts, J. B. F. N. Hydrophobic effects—opinions and facts. *Angew. Chem. Int. Edn Engl.* **32**, 1545–1579 (1993).
- Kunitake, T. Synthetic bilayer-membranes—molecular design, self-organization, and application. *Angew. Chem. Int. Edn Engl.* **31**, 709–726 (1992).
- Beijer, F. H., Kooijman, H., Spek, A. L., Sijbesma, R. P. & Meijer, E. W. Self-complementarity achieved through quadruple hydrogen-bonding. *Angew. Chem. Int. Edn Engl.* **37**, 75–78 (1998).
- Lydon, J. Chromonic liquid crystal phases. *Curr. Opin. Colloid Interface Sci.* **3**, 458–466 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank K. Pieterse for creation of the molecular cartoons; Laboratoire Léon Brillouin for providing the SANS beam time; and F. Boué for assistance. This work was supported by The National Research School Combination: Catalysis and the Dutch Science Foundation (NWO); we thank DSM Research for a grant.

Correspondence and requests for materials should be addressed to E.W.M. (e-mail: E.W.Meijer@tue.nl).

Termination of global warmth at the Palaeocene/Eocene boundary through productivity feedback

Santo Bains*, Richard D. Norris†, Richard M. Corfield* & Kristina L. Faul‡

* Department of Earth Sciences, Parks Road, University of Oxford, Oxford OX1 3PR, UK

† Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02540-1541, USA

‡ Earth Sciences Department, University of California, Santa Cruz, California 95064, USA

The onset of the Palaeocene/Eocene thermal maximum (about 55 Myr ago) was marked by global surface temperatures warming by 5–7 °C over approximately 30,000 yr (ref. 1), probably because of enhanced mantle outgassing^{2,3} and the pulsed release of ~1,500 gigatonnes of methane carbon from decomposing gas-hydrate reservoirs^{4–7}. The aftermath of this rapid, intense and global warming event may be the best example in the geological record of the response of the Earth to high atmospheric carbon dioxide concentrations and high temperatures. This response has been suggested to include an intensified flux of organic carbon from the ocean surface to the deep ocean and its subsequent burial through biogeochemical feedback mechanisms⁸. Here we present firm evidence for this view from two ocean drilling cores, which record the largest accumulation rates of biogenic barium—indicative of export palaeoproductivity—at times of maximum global temperatures and peak excursion values of $\delta^{13}\text{C}$. The unusually rapid return of $\delta^{13}\text{C}$ to values similar to those before the methane release⁷ and the apparent coupling of the accumulation rates of biogenic barium to temperature, suggests that the enhanced deposition of organic matter to the deep sea may have efficiently cooled this greenhouse climate by the rapid removal of excess carbon dioxide from the atmosphere.

Mean surface air temperatures at the end of the Palaeocene epoch were ~6.7–8.5 °C warmer than today⁹. Despite the already warm temperatures, this period in the Earth's history was punctuated by a ~60,000-year episode of intense global warming at the Palaeocene/Eocene (P/E) boundary^{1,4,10}. It is likely that this warmth was caused primarily by increased atmospheric CO₂ concentrations released by tectonic events^{2,3} and was related to the destabilization (and oxidation) of sea-floor gas-hydrate accumulations^{4–7}. Gas hydrates are naturally occurring solids composed of mainly water and methane molecules, and are found in geographically diverse continental margin settings¹¹. Low temperatures (< 7 °C) and high pressures (> 50 bar) maintain the crystalline lattices¹¹ of these hydrates, which are present in vast quantities today (of the order of 10⁴ gigatonnes (Gt) of methane carbon, or double the estimate of carbon from all known fossil-fuel sources)¹¹. Methane from gas hydrates has an average $\delta^{13}\text{C}$ of –60‰ (ref. 11), and its pulsed release at the P/E boundary is shown by a very marked, stepped, negative shift in the global $\delta^{13}\text{C}$ record that accompanied the rapid warming^{4–7}. On a millennial timescale, the rate of greenhouse-gas emission at the P/E boundary is approximately comparable to current industrial levels, and thus provides a unique insight into our uncertain greenhouse future. In response to the changing conditions of the P/E boundary, marine and terrestrial organisms underwent a period of intense reorganization^{12,13} including a notable extinction of about 50% of the species of benthic foraminifera in the deep sea¹⁴.

The oceans contain about 60 times more carbon than the atmosphere, and it has been suggested that the CO₂ concentration of the atmosphere can be controlled by changes in the export productivity

of the oceans¹⁵. During photosynthesis, marine phytoplankton fix CO₂ into their tissues; these tissues eventually settle to the deep sea (where they become part of the sedimentary organic matter) or are oxidized in transit¹⁶. The ocean surface water, depleted in CO₂, then re-equilibrates with the atmosphere^{15,16}. The availability of nutrients (particularly P, N and Fe) in surface waters is often the limiting factor for primary productivity^{17,18}. An increased input of nutrients—by, for example, upwelling from the deep sea—can cause blooms of phytoplankton, and increased export of carbon into the ocean interior: this is the process known as the ‘carbon pump’. Here we present evidence of enhanced organic-matter production and burial in the oceans (via proposed biogeochemical feedback mechanisms⁸) following gas-hydrate dissociation at the P/E boundary, and suggest that this may have led to rapidly reduced global temperatures. A combination of optimal conditions (increased runoff from the continents¹⁹, oceanic fertilization from volcanic fallout², rising global temperatures^{1,4,20} and increased atmospheric CO₂ concentrations^{2–7}) during the P/E thermal maximum seems to have triggered a bloom in marine phytoplankton, which may have sequestered the greenhouse gas CO₂ to the deep sea by about 60,000 years of enhanced biological ‘pumping’⁸.

Biogenic barium (Ba_{biogenic}) is well suited for use as a palaeoproductivity proxy. In the modern open ocean, the flux of Ba_{biogenic} correlates well with export production (that is, organic-carbon flux from surface waters) in both sediment traps²¹ and in deep-sea sediments²². In sediments, barite—the main carrier of Ba related to biological productivity²¹—is well preserved in pelagic settings and has a burial efficiency of 30% (refs 21, 23); it is not affected by diagenesis in oxic environments²⁴. The refractory nature of barite is particularly useful for studying the P/E thermal maximum because

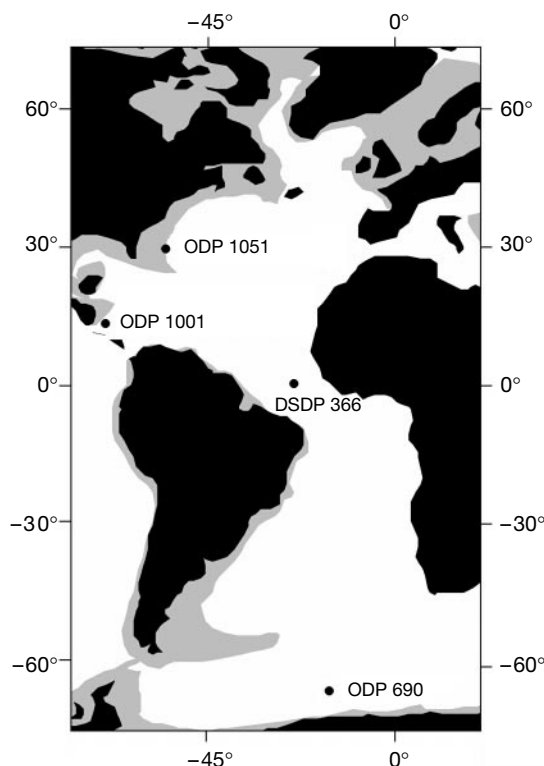


Figure 1 Palaeocene location of the sites analysed for biogenic-barium concentrations across the P/E boundary. DSDP 366: Sierra Leone rise, mid-Atlantic Ocean (present location: 05° 41' N, 19° 51' W). ODP 690: Maud rise, Weddell Sea, Antarctica (present location: 65° 09' S, 01° 12' E). ODP 1001: lower Nicaraguan rise, Caribbean Sea (present location: 15° 45' N, 74° 55' W). ODP 1051: Blake nose, western North Atlantic Ocean (present location: 30° 03' N, 76° 21' W). Shaded areas indicate ocean depths less than 500 m below sea level.

other proxies (for example, CaCO_3 , biogenic silica and organic carbon) are subject to significant remineralization and cannot comprehensively define the nature of productivity changes over this interval. Barite is thought to form in the upper water column²⁵, in microenvironments associated with decomposing organic matter

such as the remains of siliceous plankton^{26,27} and acantharian shells²⁷, or possibly from direct precipitation by living biological organisms²⁸. Although the exact mechanism of barite formation remains a topic of debate, there is abundant evidence for a strong relationship between barite and export productivity in the modern ocean. Accordingly, in the palaeorecord, high accumulation rates of $\text{Ba}_{\text{biogenic}}$ in deep-sea drill cores are thought to correspond to episodes of increased export palaeoproductivity, and have thus been used as export productivity tracers throughout the Cenozoic^{22,29}. For the purposes of our study, we calculate the flux of $\text{Ba}_{\text{biogenic}}$ and account for Ba from detrital sources by normalizing to Al (according to the method of Dymond *et al.*²¹).

We have determined the $\text{Ba}_{\text{biogenic}}$ concentrations of 130 marine samples across the P/E boundary, from two geographically diverse Ocean Drilling Program (ODP) sites 690 and 1051 (Fig. 1), and have converted these values into mass accumulation rates (MARs) using a well constrained high-resolution orbital timescale based on precessional cycles¹⁰. Our results clearly show that the decline of the global greenhouse climate coincided with an accelerated $\text{Ba}_{\text{biogenic}}$ flux at both ODP sites (Fig. 2). The calculated accumulation rates return to pre-excursion levels in a manner virtually identical to that of published $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ records^{1,4}.

Although $\text{Ba}_{\text{biogenic}}$ cannot be used to record accurately changes in export productivity under highly reducing conditions (because Ba remobilization may occur due to sulphate reduction), barite is well preserved in oxic sediments such as those studied here²⁴. Additionally, Ba remobilization results in a negative inflection in $\text{Ba}_{\text{biogenic}}$, whereas the excursion that we observe is positive. Furthermore, the similarity of the $\text{Ba}_{\text{biogenic}}$ records and published $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$

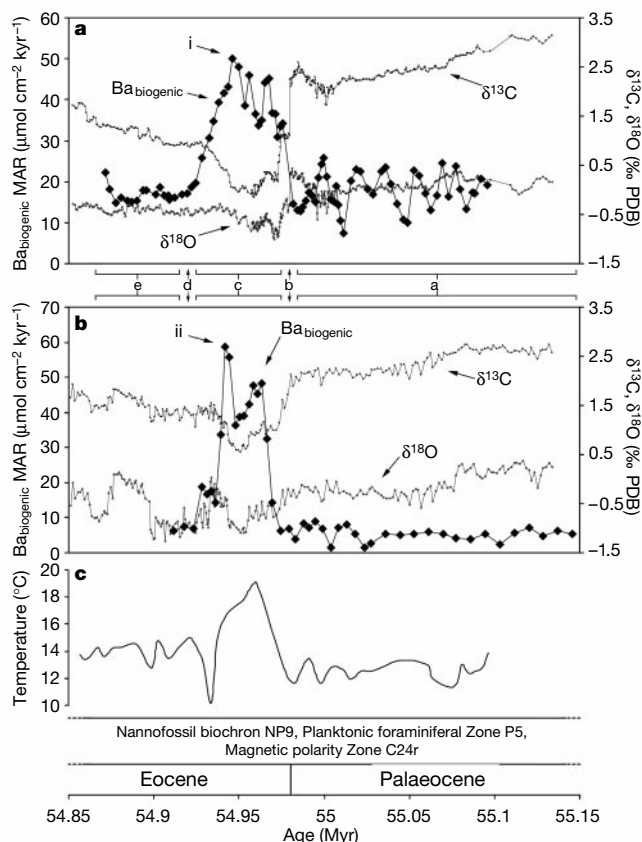


Figure 2 Variation across the P/E boundary of the properties of two deep-sea sediment cores. The figure shows biogenic-barium mass accumulation rate ($\text{Ba}_{\text{biogenic}}$ MAR), and $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ records⁴ from ODP sites 690B (**a,c**) and 1051B (**b**). The Antarctic Ocean water temperature change ($^{\circ}\text{C}$) throughout this interval closely resembles inferred export productivity variations at both sites (**c**). The temperature stratigraphy was generated from published benthic foraminifer *Nuttallides truempyi* $\delta^{18}\text{O}$ data¹. Before the P/E boundary, the $\text{Ba}_{\text{biogenic}}$ flux remains relatively stable at sites 690B and 1051B (feature a). Following the initial burst of warmth and carbon-isotope excursion, $\text{Ba}_{\text{biogenic}}$ MARs dramatically increase in both the mid- and high-latitude Atlantic Ocean (feature b). $\text{Ba}_{\text{biogenic}}$ MARs rise from 6 to $48 \mu\text{mol Ba cm}^{-2} \text{ kyr}^{-1}$ in ODP 1051B, and from 14 to $34 \mu\text{mol Ba cm}^{-2} \text{ kyr}^{-1}$ in ODP 690B within $\sim 5,000\text{--}7,000$ years. Subsequently, for an interval of about 60,000 years (corresponding to the warmest interval of the P/E transition; feature c), $\text{Ba}_{\text{biogenic}}$ MARs remain high with peak flux of $59 \mu\text{mol Ba cm}^{-2} \text{ kyr}^{-1}$ reached in the western North Atlantic Site 1051B (ii) and $50 \mu\text{mol Ba cm}^{-2} \text{ kyr}^{-1}$ in the Antarctic Site 690B (i). Global temperatures trends vary with $\text{Ba}_{\text{biogenic}}$ MARs as the records return to near pre-excursion levels (feature d). The fact that $\text{Ba}_{\text{biogenic}}$ MARs stabilize at the same time in two geographically diverse cores strongly suggests that the mechanism that was driving enhanced export productivity became exhausted at all latitudes simultaneously. In the aftermath of this intense global warming (and cooling) anomaly, $\text{Ba}_{\text{biogenic}}$ MARs and global temperatures remain relatively stable into the early Eocene epoch (feature e). The fact that maximum $\text{Ba}_{\text{biogenic}}$ MARs are reached first in site 690B (i) could reflect the high-latitude oceans being more rapidly affected by changing atmospheric greenhouse-gas concentrations. Recent concerns that some Al in oceanic sediments might also be of biogenic origin³² prompted us to verify the fidelity of our results by normalization using Ti. $\text{Ba}_{\text{biogenic}}$ concentrations obtained by this approach were virtually identical to those obtained using Al, and Al:Ti ratios indicated no Al from biogenic sources. All elemental abundance data were generated by inductively coupled plasma atomic emission spectrometry (ICP-AES) analysis of crushed bulk sediments dissolved in a 1:2 mixture of HClO_4 and HF acids, then diluted with HCl and distilled water.

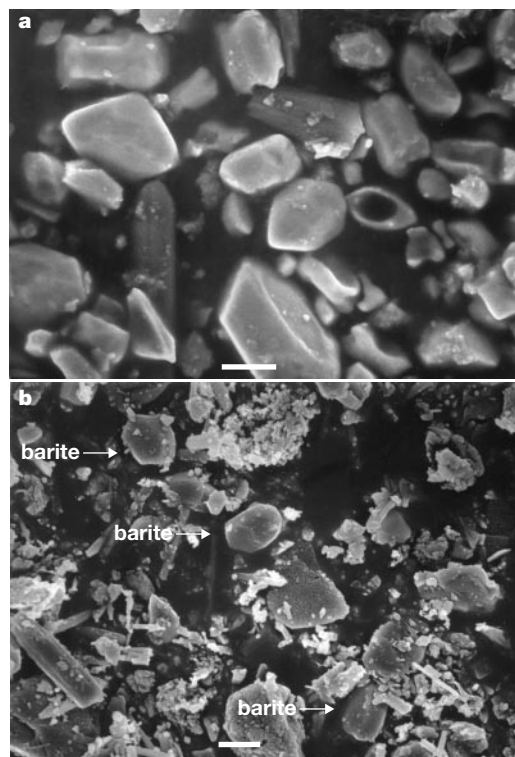


Figure 3 Scanning electron micrographs of marine barite microcrystals ($1\text{--}5 \mu\text{m}$ in size) from samples during (**b**) and after (**a**) the P/E boundary. **a**, Sediment significantly above the P/E boundary from Site 1051A (171B 1051 A, 48X-4, 71–73 cm, depth 453 m below sea floor, m.b.s.f.); **b**, sediment from Site 1051B where $\text{Ba}_{\text{biogenic}}$ MAR values rapidly increase following the primary $\delta^{13}\text{C}$ excursion at the P/E boundary, 512.75 m.b.s.f., approximately 54.966 Myr ago. Sample **b** was used in this study, and in addition to the barite crystals, dissolution-resistant terrigenous particles are present, including titanium oxides and aluminosilicates. Each scale bar is $\sim 2 \mu\text{m}$.

data^{1,4} suggest that all are largely controlled by perturbations in the carbon cycle and global climate at the P/E boundary. We believe, therefore, that the Ba_{biogenic} MAR anomaly is not a diagenetic phenomenon. Diagenesis would be expected to strip Ba out of pore waters, and cause the baseline values following the excursion to be lower than before the excursion³⁰. Our data from both sites rise from, and fall to, the same baseline values. Furthermore, we would expect Ba_{biogenic} MARs to begin to increase below the onset of the $\delta^{13}C$ anomaly if the Ba_{biogenic} MAR peak were due entirely to the diffusion of sulphate from underlying oxic sediments. To confirm that our samples contain biogenic barite, we separated marine biogenic barite in selected samples using the method of Paytan *et al.*²⁴, and observed distinct ovoid barite crystals using scanning electron microscopy and elemental analysis (SEM/EDS) in samples from site 1051B taken from the P/E thermal maximum interval, and from site 1051A slightly above the interval (Fig. 3). Barite crystals with this type of morphology have been identified in sediments underlying other ancient high-productivity regions²⁴. The exceptional preservation of these crystals in both samples strongly suggests that our Ba_{biogenic} MAR findings reflect ancient surface export production, and that neither Ba remobilization nor diagenesis has significantly affected our records.

Recent analysis using mass-balance equations suggests that the recovery of the $\delta^{13}C$ record at ODP site 1051 occurred more rapidly than would be expected through a steady-state carbon cycle⁷. The most plausible explanation for this would be a global increase in carbon throughput and organic-carbon burial following gas-hydrate dissociation⁷. The similarity in shape of the $\delta^{13}C$ and Ba_{biogenic} records at both sites (an abrupt excursion followed by a gradual return to base levels) indicates that Ba_{biogenic} can indeed be used as an export productivity proxy, and that our findings are consistent with the operation of a globally enhanced biological 'pump' that preferentially sequestered large quantities of ^{12}C to the ocean depths, stimulated by changes to the global climate and carbon cycle during the P/E thermal maximum.

Both ODP sites used in our study are near continental land-masses, suggesting that our data could reflect an export productivity and burial phenomenon at depositional environments of this sort world-wide, as the Earth's carbon cycle laboured to reach equilibrium following massive input of CO_2 to the atmosphere. A strong productivity signal in eastern Tethys during the late Palaeocene, thought to be due primarily to vigorous upwelling²⁹, may also have been caused by the mechanism we report here. In further support of the global scale of the biological response that we discuss, preliminary work has shown elevated Ba concentrations in Caribbean (ODP 1001) and eastern tropical Atlantic (DSDP 366) sediments that contain the P/E boundary.

Some previous studies of P/E boundary sections have suggested decreased productivity^{13,20}. Nonetheless, in the modern ocean, regional oceanographic conditions can vary significantly, and it is not unlikely that there was an overall increase in carbon burial even if some sites became oligotrophic. The structure of complete $\delta^{13}C$ records from the P/E boundary (ref. 4), and their faster-than-expected recovery to pre-excursion levels⁷ (coupled with our Ba_{biogenic} records), transcend regional variations, and indicate an increase in the global-average organic matter export and burial.

It is probable that much of the organic matter deposition associated with the P/E boundary occurred in near-shore marine environments⁸. As the climate became very humid and wet, increased rainfall could have caused enhanced erosion of the continents (from chemical weathering), and may have led to a boost in the nutrient supply to the oceans from terrestrial runoff⁸. This, coupled with increased temperatures and atmospheric CO_2 concentrations, and fertilization by volcanic fallout, could have caused blooms in many oceanic regions. We also note that, although the ocean is the largest sink for atmospheric CO_2 , P/E boundary conditions were probably also favourable for the

increased abundance of terrestrial biota (particularly plants) and carbon-bearing soils, which almost certainly would have further accelerated the recovery of the global carbon cycle as geographic coverage increased^{8,31}.

We have shown the existence of an export productivity maximum at the P/E boundary, lending support to the theory that these conditions produced a series of negative feedbacks⁸ in the climate system. These feedbacks acted as an immense 'biological pump', effectively reducing the increased atmospheric concentration of greenhouse gases and rapidly returning the Earth to average late Palaeocene conditions. It is possible that biological activity of this type might eventually reduce the atmospheric concentrations of CO_2 that are at present rising due to anthropogenic emissions. However, we note that it took ~60,000 years for late Palaeocene temperatures to return to normal following the apparently cataclysmic decomposition of gas hydrates. Furthermore, the current potential of the 'biological pump' may have been reduced by deforestation and pollution of the oceans. □

Received 6 March; accepted 17 July 2000.

- Kennett, J. P. & Stott, L. D. Abrupt deep-sea warming, palaeoceanographic changes and benthic extinctions at the end of the Palaeocene. *Nature* **353**, 225–229 (1991).
- Eldholm, O. & Thomas, E. Environmental impact of volcanic margin formation. *Earth Planet. Sci. Lett.* **117**, 319–329 (1993).
- Rea, D. K., Zachos, J. C., Owen, R. M. & Gingerich, P. D. Global change at the Paleocene-Eocene boundary: climatic and evolutionary consequences of tectonic events. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **79**, 117–128 (1990).
- Bains, S., Corfield, R. M. & Norris, R. D. Mechanisms of climate warming at the end of the Paleocene. *Science* **285**, 724–727 (1999).
- Dickens, G. R., O'Neil, J. R., Rea, D. K. & Owen, R. M. Dissociation of oceanic methane hydrate as a cause of the carbon isotope excursion at the end of the Paleocene. *Paleoceanography* **10**, 965–971 (1995).
- Dickens, G. R., Castillo, M. M. & Walker, J. C. G. A blast of gas in the latest Paleocene: Simulating first-order effects of massive dissociation of oceanic methane hydrate. *Geology* **25**, 259–262 (1997).
- Dickens, G. R. Carbon addition and removal during the Late Paleocene Thermal Maximum: Basic theory with a preliminary treatment of the isotope record at Ocean Drilling Program Site 1051, Blake Nose. In *Western North Atlantic Paleogene and Cretaceous Paleoclimatology* (eds Kroon, D., Norris, R. D. & Klaus, A.) (Special Publication, Geological Society, London, in the press).
- Zachos, J. C. & Dickens, G. R. An assessment of the biogeochemical feedback response to the climatic and chemical perturbations of the LPTM. *GFF* **122**, 188–189 (2000).
- Budyko, M. Climate catastrophes. *Glob. Planet. Change* **20**, 281–288 (1999).
- Norris, R. D. & Röhl, U. Carbon cycling and chronology of climate warming during the Paleocene/Eocene transition. *Nature* **401**, 775–778 (1999).
- Kvenvolden, K. A. In *Gas Hydrates* (eds Henriot, J.-P. & Mienert, J.) 9–30 (Geological Society, London, 1998).
- Koch, P. L., Zachos, J. C. & Gingerich, P. D. Correlation between isotope records in marine and continental carbon reservoirs near the Paleocene/Eocene boundary. *Nature* **358**, 319–322 (1992).
- Kelly, D. C., Bralower, T. J. & Zachos, J. C. Evolutionary consequences of the latest Paleocene thermal maximum for tropical planktonic foraminifera. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **141**, 139–161 (1998).
- Pak, D. K. & Miller, K. G. Late Paleocene to early Eocene benthic foraminiferal stable isotopes and assemblages: implications for deepwater circulation. *Paleoceanography* **7**, 405–422 (1992).
- Broecker, W. S. Glacial to interglacial changes in ocean chemistry. *Prog. Oceanogr.* **11**, 151–197 (1982).
- Corfield, R. M. Paleocene oceans and climate: An isotopic perspective. *Earth-Sci. Rev.* **37**, 225–252 (1994).
- Tyrrell, T. The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* **400**, 525–531 (1999).
- Falkowski, P. G., Barber, R. T. & Smetacek, V. Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**, 200–206 (1998).
- Robert, C. & Kennett, J. P. Antarctic subtropical humid episode at the Paleocene-Eocene boundary: clay-mineral evidence. *Geology* **22**, 211–214 (1994).
- Thomas, D. J., Bralower, T. J. & Zachos, J. C. New evidence for subtropical warming during the late Paleocene thermal maximum: Stable isotopes from Deep Sea Drilling Project Site 527, Walvis Ridge. *Paleoceanography* **14**, 561–570 (1999).
- Dymond, J., Suess, E. & Lyle, M. Barium in deep-sea sediment: A geochemical proxy for productivity. *Paleoceanography* **7**, 163–181 (1992).
- Paytan, A., Kastner, M. & Chavez, F. P. Glacial to interglacial fluctuations in productivity in the Equatorial Pacific as indicated by marine barite. *Science* **274**, 1355–1357 (1996).
- Paytan, A. & Kastner, M. Benthic Ba fluxes in the central Equatorial Pacific, implications for the oceanic Ba cycle. *Earth Planet. Sci. Lett.* **142**, 439–450 (1996).
- Paytan, A., Kastner, M., Martin, E. E., Macdougall, J. D. & Herbert, T. Marine barite as a monitor of seawater strontium isotope composition. *Nature* **366**, 445–449 (1993).
- Legelux, F. & Reys, J. $^{226}Ra/^{228}Ra$ activity ratio in oceanic settling particles: implications regarding the use of barium as a proxy for paleoproductivity reconstruction. *Deep-Sea Res.* **143**, 1857–1863 (1996).
- Bishop, J. K. B. The barite-opal-organic carbon association in oceanic particulate matter. *Nature* **332**, 341–343 (1988).
- Bernstein, R. E., Byrne, R. H. & Schijf, J. Acantharians: a missing link in the oceanic biogeochemistry of barium. *Deep-Sea Res.* **145**, 491–505 (1998).

28. Bertram, M. A. & Cowen, J. P. Morphological and compositional evidence for biotic precipitation of marine barite. *J. Mar. Res.* **55**, 577–593 (1997).
29. Schmitz, B., Charisi, S. D., Thompson, E. I. & Speijer, R. P. Barium, SiO₂ (excess), and P₂O₅ as proxies of biological productivity in the Middle East during the Palaeocene and the latest Palaeocene benthic extinction event. *Terra Nova* **9**, 95–99 (1997).
30. McManus, J. *et al.* Geochemistry of barium in marine sediments: Implications for its use as a paleoproxy. *Geochim. Cosmochim. Acta* **62**, 3453–3473 (1998).
31. Beerling, D. J. The responses of the terrestrial carbon cycle to global change across the Paleocene/Eocene boundary. *GFF* **122**, 23 (2000).
32. Murray, R. W., Leinen, M. & Isern, A. R. Biogenic flux of Al to sediment in the central Equatorial Pacific Ocean: Evidence for increased productivity during glacial periods. *Paleoceanography* **8**, 651–670 (1993).

Acknowledgements

We thank N. Paige, S. James, N. Walsh, S. Wyatt and J. Cartledge for their technical assistance; and G. Dickens, M. Delaney and A. Paytan for discussions and comments. We also thank A. Paytan for use of her laboratory to separate barite. Samples for this study were taken from cores drilled and supplied by the Ocean Drilling Program. The Natural Environment Research Council (UK), NSF and JOI/USSP supported this work.

Correspondence and requests for data should be addressed to S.B.
(e-mail: santo.bains@earth.ox.ac.uk).

Earthquake-induced changes in a hydrothermal system on the Juan de Fuca mid-ocean ridge

H. Paul Johnson*, Michael Hutnak*, Robert P. Dziak†, Christopher G. Fox‡, Istvan Urcuyo‡, James P. Cowen§, John Nabelek|| & Charles Fisher‡

* School of Oceanography, University of Washington, Seattle, Washington 98195-7940, USA

† PMEL NOAA and Oregon State University, Newport, Oregon 97365, USA

‡ Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA

§ Department of Oceanography, University of Hawaii, Honolulu, Hawaii 96822, USA

|| COAS, Oregon State University, Corvallis, Oregon 97331, USA

Hydrothermal vents on mid-ocean ridges of the northeast Pacific Ocean are known to respond to seismic disturbances, with observed changes in vent temperature^{1–4}. But these disturbances resulted from submarine volcanic activity; until now, there have been no observations of the response of a vent system to non-magmatic, tectonic events. Here we report measurements of hydrothermal vent temperature from several vents on the Juan de Fuca ridge in June 1999, before, during and after an earthquake swarm of apparent tectonic origin. Vent fluid temperatures began to rise 4–11 days after the first earthquake. Following this initial increase, the vent temperatures oscillated for about a month before settling down to higher values. We also observed a tenfold increase in fluid output from the hydrothermal system over a period of at least 80 days, extending along the entire ridge segment. Such a large, segment-wide thermal response to relatively modest tectonic activity is surprising, and raises questions about the sources of excess heat and fluid, and the possible effect on vent biological communities.

The earthquake swarm occurred on 8 June, and was located on the Endeavour segment of the Juan de Fuca ridge (Fig. 1a), a well studied spreading centre with extensive hydrothermal activity (Fig. 1b; ref. 5), but one that lacks evidence of recent volcanic activity. Unlike the southern Juan de Fuca ridge, the Endeavour

segment is very active seismically, producing hundreds of small (magnitude $M < 4$) earthquakes per year. Seismicity from submarine volcanic events is characterized by steady, low-level earthquakes, horizontal migration of epicentres due to lateral dyke injection, and continuous, low-frequency volcanic tremor. In contrast, earthquakes generated by non-magmatic, tectonic movements along normal or strike-slip faults are also frequent in the northeast Pacific, and are characterized by a large initial main shock, followed by aftershocks that decay quasi-exponentially through time with no associated volcanic tremor. The 8 June earthquake is interpreted to be tectonic in origin because the initial earthquake was the largest of the swarm, there was quasi-exponential decay in the number of aftershocks, and volcanic tremor was absent. This event was significantly more intense than similar tectonic sequences observed at the Endeavour segment, with a main shock of magnitude 4.5 and an unusually large number of aftershocks occurring during five days of activity (Fig. 1a). The main shock was recorded by the Pacific Northwest regional seismic network, and the focal mechanism was determined using moment tensor inversion techniques⁶ to be the result of slip on a normal fault striking 160°, further supporting the tectonic origin of this event.

That the 8 June event was 'non-eruptive' is strongly suggested by data obtained during extensive dives by the submersibles *Alvin* and *Jason* during August–September 1999 in the Endeavour axial valley: no evidence was found of any new volcanic activity in multiple vehicle traverses between the Beach and Clam Bed sites (see Fig. 1 for locations). However, a great increase of particulate matter in the bottom water was observed at all sites during these dives, compared to previous years. A ubiquitous white flocculent material was suspended in the near-bottom water and covered much of the areas of diffuse flow; several diffuse vents were still emitting this material in August–September. Microscopic examination confirmed that the flocculent material was largely inorganic, presumably sulphur of biogenic origin (D.C. Nelson, personal communication), similar to that reported coming from 'snow-blower vents', which accompany crustal disturbances associated with submarine eruptions^{7,8}. Most of the 8 June aftershock seismicity had ceased by 13 June.

The earthquake locations shown in Fig. 1a indicate that the locus of seismic activity is offset to the west of the Endeavour axial valley, with a mean event location (weighted by epicentral uncertainty) of 47° 52.4' N and 129° 14.2' W. A bias in epicentre locations is possible due to sound velocity or bathymetry effects, but the distribution of locations indicate that the swarm was probably centred 7.5 km west of the segment axis. This area of the Endeavour western flank has large, exposed, inward-facing fault scarps associated with crustal stretching that have been observed during previous *Alvin* dives⁹.

During the 10-month interval preceding 8 June, hydrothermal vent temperatures on the Endeavour segment were monitored at the Easter Island site (EI/MEF, within the main Endeavour field) and the Clam Bed site (650 m south of the High Rise field). At these locations, water temperatures associated with *vestimentiferan* tube-worm aggregations around diffuse vents were monitored using four pairs of thermistors, with the lower (near-bottom) sensor of each pair embedded among the bases of the tubeworms within diffuse flow, and the upper sensor near the plumes of the tubeworms, approximately 0.5 m above the sea floor. The Beach site (300 m south of the MEF) is a small sediment pond within the axis, with diffuse hydrothermal fluid venting around and through the sediments. At this site, a bio-column installation was filtering hydrothermal fluids to isolate sub-surface microbial populations, and the hydrothermal effluent was contained entirely within the plumbing of the enclosed bio-filter system. The hydrothermal fluid temperatures were measured within a PVC filter shell, isolated from normal bottom-water variations. The bio-column data logger also had a companion external thermistor for measuring the general

bottom-water temperatures. Recovery of the data loggers at all three sites took place in August–September, 1999.

Before 8 June, data from all sites indicate that vent temperatures were very stable for the previous 10 months, with only a semi-diurnal tidal modulation providing a small ($< 5\%$) variation in fluid temperatures. All three sites were located within diffuse vent fields, and all had temperatures before 8 June that were only slightly above normal bottom-water values. Before 8 June, none of the sites showed the oscillations that were present after the earthquake event, indicating that the lower thermistors were deployed in a stable vent effluent. The Beach and Clam Bed sites are isolated regions of low-temperature diffuse vents separated from the adjacent large high-temperature fields by several hundred metres of volcanic sea floor with no known hydrothermal activity (Fig. 1b). In contrast, the EI/MEF site is a diffuse flow area within the large-scale main Endeavour field, and the thermistors were within 10 m of the ‘Peanut’ sulphide edifice¹⁰.

Exit fluid temperatures at the monitored sites began to increase 4 to 11 d after the onset of seismicity, similar to the delay observed at high-temperature vents on the East Pacific Rise¹¹. This delay may be a characteristic time constant associated with the reaction of the crustal fluid flow regime to a disturbance. After delays of several days, all thermistor records show a systematic increase in mean temperature of the hydrothermal vent fluids. Vent water temperatures first increased at the EI/MEF site (Fig. 2b), where the monitored vents were directly adjacent to the high-temperature smokers⁵. Because tidal modulation and the slow rate of temperature increase make it difficult to identify the ‘first arrival time’ of the disturbance, this can be only estimated at EI/MEF as 4 ± 1 d after 8 June. At Clam Bed and Beach, ‘first arrival’ disturbances were more abrupt, occurring almost 11 d (Fig. 2) after 8 June. Although

there is some variability in ‘first arrival times’ for the four thermistors at the Clam Bed site, the 11-d mean time for the initiation of the change is similar to the Beach site. This temporal coincidence is unexpected, because the Clam Bed and Beach sites are separated by 2 km, and each site is adjacent to a different, large, high-temperature hydrothermal field (Fig. 1b) with presumably distinct sub-surface circulation systems⁵.

A remarkable post-event feature for all three sites was the appearance of large-amplitude temperature oscillations after the maximum disturbed temperature was reached (Fig. 2). In some cases, these temperature oscillations were $\pm 5^\circ\text{C}$, with periods between 8 and 12 d. These high-amplitude variations appeared correlated within each individual site, but poorly correlated between sites (Fig. 2). After several oscillation periods, the variability at the EI/MEF site decreased, but the vent temperatures were still substantially higher than the pre-event mean values when the monitoring systems were recovered. The Beach data logger ran out of memory on 24 July 1999.

The physical processes responsible for the high-amplitude temperature oscillations, beginning more than 20 d after the initial tectonic event, are not clear. These oscillations could represent the non-uniform advance of a ‘cracking front’ in the lower crustal rocks, with each episodic propagation of cracks into a region of new hot rock producing temperature increases in the hydrothermal system. In this explanation, the short-term fluid temperature ‘spikes’ would have a relatively short duration, as heat would be effectively transferred to the upwelling hydrothermal fluid within the small, newly cracked zone. Another possible mechanism to produce the oscillations would be the episodic plugging and unblocking of the hydrothermal sub-surface ‘plumbing’ by biologically generated material, specifically the abundant flocculent material that was

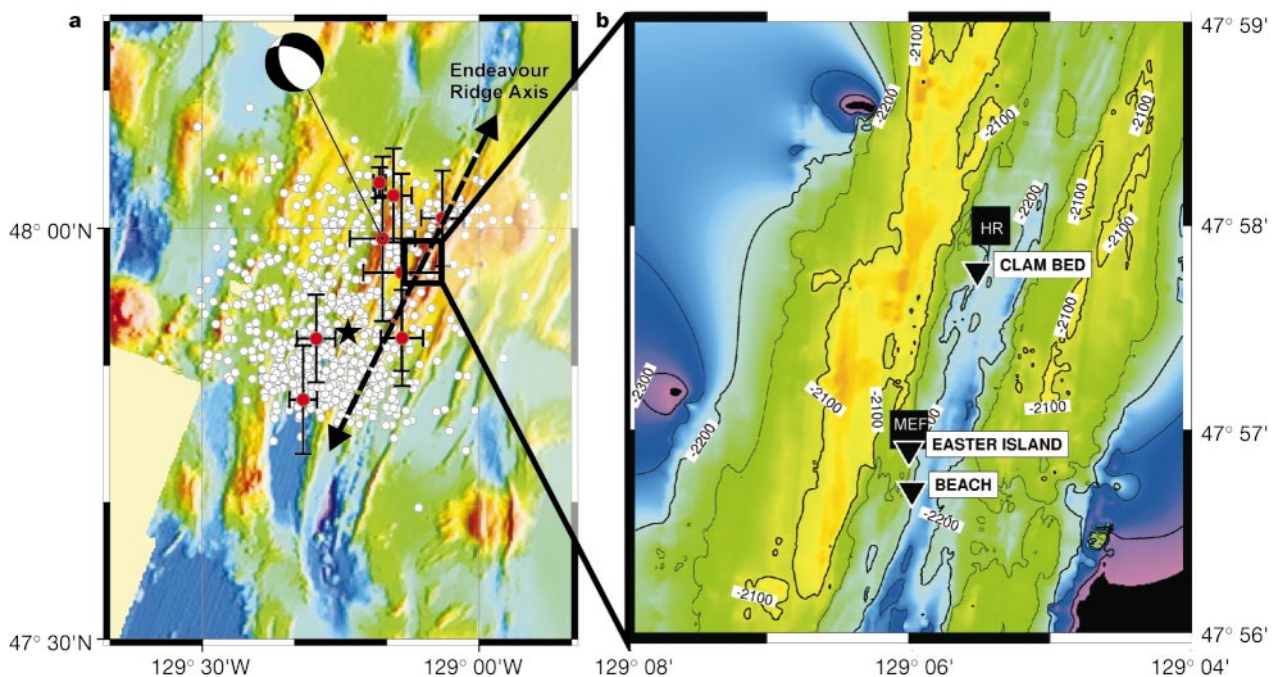


Figure 1 Bathymetry of the Endeavour segment of the Juan de Fuca ridge. The figure shows locations of earthquakes (**a**) and hydrothermal vent monitors (**b**). In **a**, the epicentres of 747 earthquakes (of the total 2,687 detected) from the June 1999 Endeavour swarm are shown as white circles. Red circles represent the eight largest events detected, with the brackets showing their location error at the 95% confidence interval (CI). Smaller events were less well constrained, with a mean error of $\pm 6.6'$ in latitude, and $\pm 9.9'$ in longitude, at the 95% CI. The mean position of swarm is shown by the black star. The main shock was located west of the axis, and aftershock activity was also centred southwest of the axial valley, indicating an off-axis location for most of the

earthquakes. The main shock mechanism is shown as a lower hemisphere equal area projection, with compression quadrants shaded. The axis of spreading is shown as the line oriented 020° N. **b**, Expanded view of the Endeavour axial valley, showing the location of the vent temperature monitoring sites (black triangles). MEF (Main Endeavour field) and HR (High Rise field) are large, high-temperature hydrothermal fields surrounded by major sulphide deposits. The Easter Island (EI) site is an area of diffuse venting that lies within the MEF; the Clam Bed (CB) and Beach sites are areas of low-temperature, diffuse venting isolated from the large fields.

released by the increased flow after the seismic event. Last, the observed temperature variations may simply represent a temporary unstable convection mode as the fluid circulation comes to a new equilibrium after the tectonic disturbance. Our present data are inadequate to identify the specific processes responsible for this unexpected oscillatory behaviour.

The thermistors recorded only temperature changes for the vent fluids associated with the seismic event; they did not record flow rates. Assuming that the reservoir of crustal fluids is a porous medium, and that the basic geometry of the system does not change substantially during the tectonic activity, Darcy's law can be used to estimate volume discharge rates from temperature changes. For a porous media, Darcy's law is given by $Q = (k/\nu) \partial P / \partial z$, where Q is the volume flow rate, k is the dynamic permeability, ν is the fluid viscosity, and $\partial P / \partial z$ is the pressure gradient that drives the flow. If we

assume that flow is driven only by the density contrast, $\Delta\rho$, due to temperature differences between the warm fluid and sea water, we can estimate the pressure gradient as $\partial P / \partial z = \Delta\rho \cdot g$. Substitution for the pre-seismic initial (Q_i) and post-seismic final (Q_f) flow rates yields the expression $Q_f / Q_i = [\gamma_i (\rho_o - \rho_i)] / [\gamma_f (\rho_o - \rho_i)]$, where subscripts i, f and o indicate initial (before 8 June), final (after 8 June) and original bottom-water parameters, respectively. The viscosity and density of sea water are well-known functions of temperature, and bottom-water and vent-fluid temperatures before 8 June are determined from the thermistor records. The resulting calculated increases in volume flux (Q_f / Q_i) from this model are 12, 4–8 and 11–21 for the Beach, EI/MEF and Clam Bed sites, respectively.

The above calculations assume a constant permeability and that additional heat is added to the system from lower crustal rocks after the earthquake, producing a positive correlation between fluid flux and temperature variations. Alternatively, the earthquake may substantially alter crustal permeability and no new heat sources may be available to the hydrothermal fluid. For this situation, total thermal output of the system can be described by $H = Q T$, where H is the heat output, Q is the volume flux described previously and T is the temperature difference between the hydrothermal fluid and sea water. If the earthquake exposed no new heat sources and energy is conserved, H must be constant, and the flux rate Q and temperature T will, in this case, be inversely related. These simple considerations suggest that an observed positive correlation between flux rate and vent temperature variation after the earthquake would imply new lower crustal thermal sources; a negative correlation would be evidence for an altered crustal permeability.

Independent of any assumptions, however, it is clear from the temperature data alone that the Endeavour hydrothermal system changed substantially after the 8 June earthquake sequence. That these changes occurred over a large segment of the axial valley, in vents associated with two presumably distinct hydrothermal fields, suggests that crustal fluid circulation beneath the entire axial valley was profoundly affected by the tectonic event.

The 8 June off-axis earthquake sequence was not rare for the Endeavour segment of the Juan de Fuca ridge, but the observation that the changes in hydrothermal systems occurred over a 2-km-long segment of the ridge axis within different environments is surprising. It is not clear how often these events occur, how long the anomalous hydrothermal output will continue, what the sources of possible excess heat and fluid might be, or what effect the increased vent and bottom-water temperatures will have on the associated biological communities. However, the marked segment-wide thermal response to the relatively modest 8 June seismic event emphasizes the dynamic nature of the evolution of a mid-ocean ridge spreading centre.

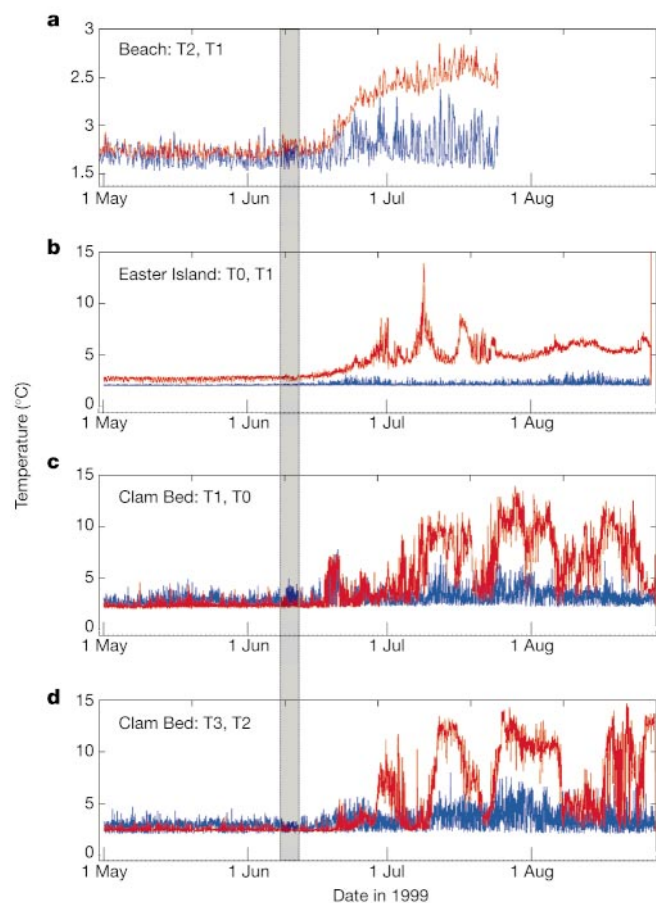


Figure 2 Temperature data from thermistors at the three Endeavour axial valley sites. The 3-month period shown includes the 8 June seismic event. Red traces, data from thermistors located within hydrothermal fluids from vents; blue traces, data from thermistors deployed in the adjacent (non-vent) bottom water at each site. Specific thermistors are labelled with T and a number. **a**, Data from the bio-column thermistors at the Beach site, located in a diffuse vent within a sediment pond 200 m south of the MEF. The earthquake activity (8–15 June 1999) is marked by the vertical shaded bar. Elevation of vent temperatures associated with the 8 June event is clearly visible in the internal thermistor, as is the increase in amplitude of the bottom-water temperature (external) variations. We note that the temperature scale for this site is different from the other three sites shown. **b**, Temperatures from the EI/MEF site, showing the slow temperature rise after the 8 June event. Oscillations begin almost 20 d after the seismic activity. **c**, Data from one of the Clam Bed thermistor pairs, near the High Rise vent field. **d**, Temperature data from a second pair of thermistors from the Clam Bed site, located only a few tens of metres from those in **c**. For the EI/MEF and Clam Bed sites, the data stream ended on 26 August 1999 and 28 August 1999, respectively; at Beach, the time series ended on 24 July 1999.

Submitted 13 December 1999; accepted 21 July 2000.

1. Fox, C. G., Dziak, R. P., Matsumoto, H. & Schreiner, A. E. Potential for monitoring low-level seismicity on the Juan de Fuca Ridge using military hydrophone arrays. *Mar. Technol. Soc. J.* **27**, 22–30 (1994).
2. Fox, C. G. *et al.* Acoustic detection of a seafloor spreading episode on the Juan de Fuca Ridge using military hydrophone arrays. *Geophys. Res. Lett.* **22**, 131–134 (1995).
3. Dziak, R. P. & Fox, C. G. The January 1998 earthquake swarm at axial volcano, Juan de Fuca Ridge: Hydroacoustic evidence of seafloor volcanic activity. *Geophys. Res. Lett.* **26**, 3429–3432 (1999).
4. Fox, C. G. & Dziak, R. P. Hydroacoustic detection of volcanic activity on the Gorda Ridge, February–March 1996. *Deep-Sea Res. II* **45**, 2513–2530 (1998).
5. Delaney, J. R. *et al.* The Endeavour Hydrothermal System I: Cellular circulation above an active cracking front yields large sulfide structures, fresh vent water and hyperthermophilic Archaea. *RIDGE Events* **8**, 11–19 (1997).
6. Nabelek, J. & Xia, C. Moment-tensor analysis using regional data: Application to the 25 March 1993, Scotts Mills, Oregon earthquake. *Geophys. Res. Lett.* **22**, 13–16 (1995).
7. Taylor, C. D. & Wirsén, C. O. Microbiology and ecology of filamentous sulfur formation. *Science* **277**, 1483–1485 (1997).
8. Holden, J. F., Summit, M. & Baross, J. A. Thermophilic and hyperthermophilic microorganisms in 3–30 °C hydrothermal fluids following a deep-sea volcanic eruption. *FEMS Microbiol. Ecol.* **25**, 33–41 (1998).

9. Johnson, H. P., Becker, K. & Von Herzen, R. P. Near-axis heat flow measurements on the northern Juan de Fuca Ridge: Implications for fluid circulation in oceanic crust. *Geophys. Res. Lett.* **20**, 1875–1878 (1993).
10. Delaney, J. R., Robigou, V., McDuff, R. E. & Tivey, M. K. Geology of a vigorous hydrothermal system on the Endeavour segment, Juan de Fuca Ridge. *J. Geophys. Res.* **97**, 19663–19682 (1992).
11. Sohn, R. A., Hildebrand, J. A. & Webb, S. C. A microearthquake survey of the high-temperature vent fields on the volcanically active East Pacific Rise (9° 50' N). *J. Geophys. Res.* **104**, 25367–25378 (1999).

Acknowledgements

We thank R. Lowell and A. Fisher for comments and suggestions. The field programmes and data analysis were supported by the National Science Foundation and the National Ocean and Atmospheric Administration.

Correspondence and requests for materials should be addressed to H.P.J. (e-mail: johnson@ocean.washington.edu).

Bacterial photosynthesis in surface waters of the open ocean

Z. S. Kolber*, C. L. Van Dover†, R. A. Niederman‡ & P. G. Falkowski*

* Environmental Biophysics and Molecular Ecology Program, Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey 08901-8521, USA

† Biology Department, College of William & Mary, Williamsburg, Virginia 23187, USA

‡ Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854-8082, USA

The oxidation of the global ocean by cyanobacterial oxygenic photosynthesis, about 2,100 Myr ago¹, is presumed to have limited anoxygenic bacterial photosynthesis to oceanic regions that are both anoxic and illuminated^{2,3}. The discovery of oxygen-requiring photosynthetic bacteria about 20 years ago⁴ changed this notion, indicating that anoxygenic bacterial photosynthesis could persist under oxidizing conditions. However, the distribution of aerobic photosynthetic bacteria in the world oceans, their photosynthetic competence and their relationship to oxygenic photoautotrophs on global scales are unknown. Here we report the first biophysical evidence demonstrating that aerobic bacterial photosynthesis is widespread in tropical surface waters of the eastern Pacific Ocean and in temperate coastal waters of the northwestern Atlantic. Our results indicate that these organisms account for 2–5% of the photosynthetic electron transport in the upper ocean.

Aerobic phototrophic bacteria^{4–7} are widely distributed in the α -subclass of Proteobacteria^{8–11}, indicating that they have evolved from several independent lines of purple photosynthetic bacteria. Nevertheless, these organisms are obligate aerobes, requiring molecular oxygen for growth and bacteriochlorophyll (bChl) *a* synthesis. Most of the known species are characterized by a wide variety and 'overabundance' of carotenoids^{7,11}. In most species, these pigments do not appear to transfer excitation energy to reaction centres¹². Furthermore, a relatively low cellular abundance of bChl *a* (refs 7, 11) indicates that the contribution of light to cellular energy demands is also low. Some species are potentially capable of anoxygenic photosynthesis, indicated by light-driven oxidation of the bacterial reaction centre P₈₇₀ and cytochromes^{13,14}, or light-stimulated uptake of CO₂ (ref. 15). However, phototrophy has not been confirmed in most strains, and the potential advantage of phototrophy to non-photosynthetic bacteria in natural habitats has not been unequivocally demonstrated. Their lack of photoautotrophy and requirement for O₂ has led to the proposal that aerobic phototrophs represent an evolutionary link between purple photosynthetic bacteria and aerobic heterotrophs⁸, or that they

attained a stable, non-photoautotrophic phase while retaining redundant, non-functional photosynthetic machinery. It is plausible, however, that some of these organisms may have adapted their photosynthetic apparatus to aerobic conditions and maintain a competent photosynthetic electron transport with efficient photoheterotrophic metabolism.

Our search for photosynthetic aerobic phototrophs in the open ocean was stimulated by the report of the occurrence of such organisms in a black-smoker plume above a deep-sea vent area on Juan de Fuca Ridge¹⁶. It was proposed that light emanating from hydrothermal vents could potentially provide a habitat for photosynthetic life^{17,18}. Using a newly developed infrared fast repetition rate (IRFRR) fluorometer, we searched for variable fluorescence transients as evidence of bacterial photosynthetic electron transport. We used water samples recovered from the hydrothermal vent area at 9° N with the deep-sea submersible, *Alvin*. We also surveyed a 1,000-km transect of the Pacific Ocean, and measured bacterial photosynthetic signals from waters off the coast of New Jersey.

Within the detection limits of our instrument (about 10 pg bChl per l), we failed to detect photosynthetic electron transport in black-smoker plumes and chimneys. However, we consistently observed fluorescence kinetic transients at 880–905 nm in the surface waters, indicative of bacterial photosynthetic electron transport. Representative kinetic profiles of bacterial (880 nm) and phytoplanktonic (Chl-*a*-containing; 685 nm) origin are shown in Fig 1. Both kinetic transients were excited by radiation at 470 nm. At this wavelength, bacterial reaction centres display a small effective absorption cross-section (about 60 Å² compared with 250 Å² for the Chl-*a*-containing phytoplankton), a high quantum yield of primary charge separation (0.55–0.60) estimated from the change in variable fluorescence¹⁹, a slow rate of overall photosynthetic electron transport and complete insensitivity to photo-inhibition. Assuming that the biomass of these organisms is, to first order, proportional to the amplitude of the measured fluorescence signal, we estimate a relative abundance of phototrophic bacteria about 1% of phytoplankton at 9° N.

Filtering surface seawater through a Whatman GF/F filter (approximate pore size of 0.8 µm) removed all of the 685-nm

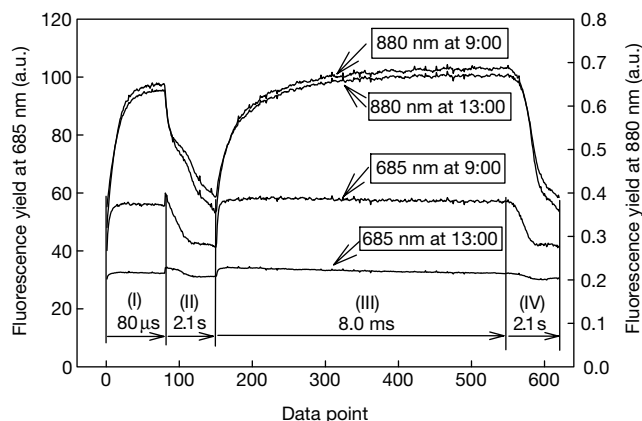


Figure 1 Typical fluorescence kinetic transients at 685 nm (planktonic photosynthesis) and 880 nm (bacterial photosynthesis). The fluorescence transients in Phase I are induced by 80 flashes of 0.4-µs duration at 1-µs intervals, resulting in reduction of the primary electron acceptor, Q_A within 20 (phytoplankton) to 60 (photosynthetic bacteria) µs. In Phase II, 70 flashes are applied at time intervals varied exponentially from 20 µs to 50 ms, resulting in a relaxation of the fluorescence signal with kinetics corresponding to the rate of electron transfer between Q_A and the quinone pool. In Phase III, 400 flashes are applied at 20–50-µs intervals causing reduction of the quinone pool. Phase IV is performed with the same protocol as in Phase II to measure the kinetics of electron transport from the quinone pool to the cytochrome *bc1* complex. The 880-nm transients displayed here are similar to those measured from laboratory cultures of *Rhodospirillum rubrum* (not shown).

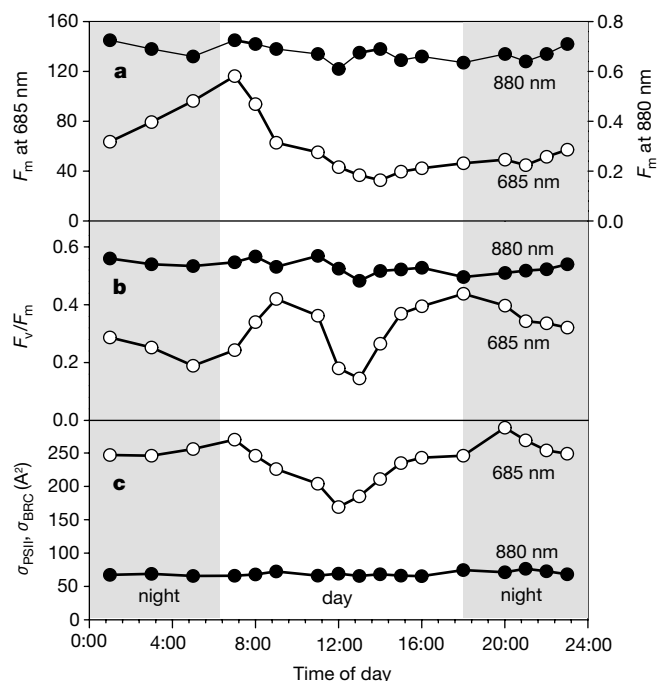


Figure 2 Diel cycle in bacterial photosynthesis (filled symbols) and planktonic photosynthesis (open symbols). **a**, Daily changes in the fluorescence yield indicating strong non-photochemical quenching in planktonic photosynthesis around local noon. **b**, Variations in F_v/F_m displaying a midday photoinhibition and midnight cell division in planktonic photosynthesis. **c**, Changes in the effective absorption cross-section. The diel cycle is absent in bacterial photosynthesis.

signal, but retained about 20% of the original 880-nm fluorescence intensity. Filtration through a 0.22- μm Nucleopore filter removed both signals. These results indicate that the majority of the phototrophic bacteria were about 0.2–0.8 μm . In samples taken off the coast of New Jersey in December 1999, only the bacterial (880-nm) fluorescence transients were observed in water passing through a 1- μm Nucleopore filter. When cultured on agar plates enriched with minimal F/2 media, the 1- μm filtrate produced individual colonies which, on transfer to liquid media, produced exclusively bacterial fluorescence transients. In all of these samples, we observed efficient excitation energy transfer from the carotenoids to bChl. The effective absorption cross-sections at 730 and 795 nm were, respectively, 15 and 3% of that at 470 nm. Our observations indicate a high concentration of carotenoid pigments, well coupled with the reaction centre antenna, but probably a paucity of either light-harvesting complexes I or II. Such a pigment organization is consistent with high light environments in the upper ocean.

Phytoplankton in the surface waters undergo a strong diel cycle in variable fluorescence. The diel pattern is manifested in variations in photochemical quantum efficiency (the F_v/F_m ratio), the maximum fluorescence yield (F_m) and the effective absorption cross-section of photosystem (PS) II (σ_{PSII} ; Fig. 2). The midday minimum in the cycle of F_v/F_m is indicative of photoinhibition or 'downregulation' of PSII photochemistry owing to overexcitation by light, leading to the formation of singlet oxygen species and damage of PSII reaction centres²⁰. The night-time minimum correlates with cell division in phytoplankton²¹ or respiratory reduction of the plastoquinone pool²². In contrast, the F_v/F_m ratio from phototrophic bacteria remained consistently high over the diel cycle, implying a lack of photoinhibition and no diel coherence in cell division (at least as reflected by photochemistry).

Using the measured photosynthetic parameters, we calculated the photosynthetically driven electron fluxes, P_e^{RC} , in both the phytoplankton and aerobic phototrophs under ambient irradiance,

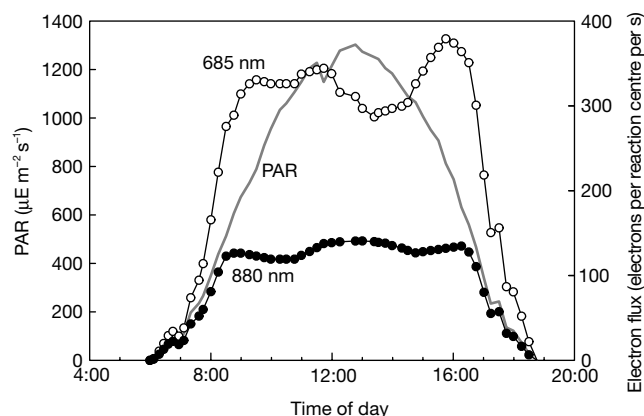


Figure 3 Reaction-centre-normalized photosynthetic electron fluxes in phytoplankton (open symbols) and aerobic phototrophs (filled symbols) calculated from Equation 1 under clear sky, and photosynthetically available radiation (PAR) conditions (continuous line). The saturation level of planktonic photosynthesis is about three times higher than in photosynthetic bacteria owing to faster electron turnover time. Also, the photosynthetic rates under subsaturating irradiance are higher in the planktonic photosynthesis owing to a much bigger effective absorption cross-section. The high saturation levels in planktonic photosynthesis are realized in the middle of the day despite strong photoinhibition (see Fig. 2a).

E , as follows²³:

$$P_e^{RC} = E \sigma_{PS} F_v/F_m \quad \text{for } E < E_s \quad (1)$$

$$P_e^{RC} = E_s \sigma_{PS} F_v/F_m \quad \text{for } E \geq E_s \quad (2)$$

where P_e^{RC} is the reaction-centre-normalized electron flux, E_s is the saturation irradiance, σ_{PS} is the effective absorption cross-section of the respective photosystem, and F_v/F_m represents the quantum yield of the primary charge separation. The results presented in Fig. 3 indicate that the photosynthetic electron fluxes in both phytoplankton and aerobic phototrophs at the surface are saturated for most of the day under clear skies, but the saturation level of bacterial electron flux is about three times lower than that of phytoplankton. The difference is partially due to the slower photosynthetic electron transport rate in aerobic phototrophs (about 11 ms per electron compared with 3 ms in phytoplankton). At subsaturating irradiances, bacterial photosynthesis is also lower as a result of the smaller effective absorption cross-section. Interestingly, the ratio of cross-section to turnover rate in phytoplankton and bacteria is highly conserved, resulting in balanced light use and electron transport in both of these organisms²⁴.

Although bacterial photosynthesis has a consistently high yield of charge separation (Fig. 2b), the overall photosynthetic efficiency of the bacterial reaction centre is about three times lower than that of the PSII/PSI reaction centre over the entire range of natural solar irradiance. However, the lower photosynthetic efficiency does not explain the much lower abundance of these organisms. We do not know the primary electron donor for these organisms, however, dissolved organic matter (DOM) is a likely source of reductant. If so, the disproportionate abundance of the phytoplankton compared with the phototrophic bacteria in the upper ocean could be due to the availability of their electron source (water), as compared with a relatively dilute pool of DOM, and the inability of these bacteria to grow photoautotrophically.

To explore the spatial relationship between bacterial and phytoplankton distribution, we measured their fluorescence signatures along a 1,000-km transect from 9.4°N to 19°N (Fig. 4a). The distribution of these two phototrophs showed a correlation over the first 200 km; however, the relative abundance of the aerobic phototrophs increased from 1% to 10% within a small distance. Chl *a* distributions, obtained from ocean colour analyses of SeaWiFS satellite images, correspond with our fluorescence-based estimates

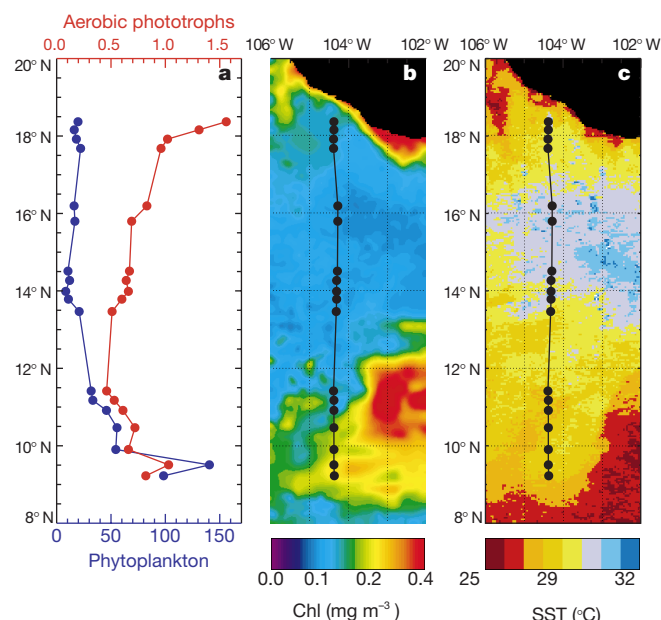


Figure 4 Spatial distribution of photosynthetic bacteria and oxygenic phytoplankton in the Pacific Ocean along 104°W in November 1999. **a**, The relative concentrations of phytoplankton (blue symbols) and aerobic phototrophs (red symbols) based on fluorescence emission signals at 685 nm and 880 nm, respectively. **b**, Spatial distribution of SeaWiFS-retrieved chlorophyll from the first week of December 1999 along the cruise track. **c**, The corresponding SST image. A relatively tight correlation between the phytoplankton and aerobic phototrophs (at about 100:1 ratio) is maintained in the high-Chl region, but the relative abundance of the aerobic phototrophs increases by a factor of ten in the low-Chl region.

of the phytoplankton concentration (Fig. 4b). The sea surface temperature (SST) data from the same period (Fig. 4c) indicate a close relationship between the physical structure of the water masses and the biological signals. Measured from natural cell populations, these signals characterize photosynthetic properties of phytoplankton/bacterial communities, each potentially composed of several different species.

Although observed in many different geographical areas, the occurrence of marine aerobic phototrophs has been assumed to be limited to small, organic-rich ecological niches such as beach sands⁴, tropical high tidal zones²⁵, mature cyanobacterial mats⁹, surfaces of green seaweeds (but not the seawater)⁴ and deep-sea hydrothermal vents (but not the surface waters directly above)²⁶. Our results indicate that phototrophic bacteria are not limited to nutrient-rich, sheltered environments, but are distributed in surface waters throughout the oceans of the world. Their abundance, relative to oxygenic phototrophs, appears to increase in oligotrophic waters, possibly reflecting the competitive advantage over non-phototrophic bacteria whose abundance is controlled solely by the concentration and quality of DOM.

The bacterial photosynthetic energy fluxes have important implications for global carbon cycling. Even though the described aerobic phototrophs may not fix inorganic carbon, the generation of ATP by photosynthesis should reduce their respiratory energy requirements in assimilating the DOM into bacterial biomass^{27,28}. This will both reduce the amount of carbon dioxide released by heterotrophic respiratory metabolism, and increase the amount of DOM assimilated into cell material. □

Methods

Measurements of photosynthetic parameters

Photosynthetic parameters of phytoplankton and photosynthetic bacteria were measured using IRFR fluorometry. This method is based on FRR¹⁹ and modified to use excitation light at 470 nm (30 nm bandwidth, 6 W cm⁻² optical power), 730 nm (50 nm bandwidth, 160 mW optical power) and 795 nm (3 nm bandwidth, 8 W cm⁻² optical power), and to measure the fluorescence transients at emission wavelengths selected from 680 nm to

950 nm using an IR-sensitive avalanche diode. The fluorescence transients were digitized at 10 MHz sampling rate and processed to calculate the effective absorption cross-section of the photosynthetic apparatus, the quantum yield of charge separation and the rates of the photosynthetic electron transport¹⁹.

Sample handling

Water samples from the hydrothermal vents site were recovered using a pair of 2 litre Niskin bottles mounted on the sample basket of the deep-sea submersible, *Alvin*. A total of 18 water samples from 9 dives were recovered and processed within 2–4 h from sample acquisition. Surface water samples were recovered using a small bucket and were processed within 10 min from sample collection. Both the raw samples (effective instrument sensitivity of 4 ng bChl per l) and filter-concentrated water samples (effective sensitivity of 10 pg bChl per l) were processed.

Received 8 May; accepted 20 June 2000.

1. Rye, R. & Holland, H. D. Paleosols and the evolution of atmospheric oxygen: a critical review. *Am. J. Sci.* **298**, 621–672 (1998).
2. Falkowski, P. G. & Raven, J. A. *Aquatic Photosynthesis* (Blackwell, Oxford, 1997).
3. Blankenship, R. E., Madigan, M. T. & Bauer, C. E. (eds). *Anoxygenic Photosynthetic Bacteria* (Kluwer, Dordrecht, 1995).
4. Shiba, T., Simidu, U. & Taga, N. Distribution of aerobic bacteria which contain bacteriochlorophyll *a*. *Appl. Environ. Microbiol.* **38**, 43–48 (1979).
5. Harashima, H., Shiba, T., Totsuka, T., Simidu, U. & Toga, N. Occurrence of bacteriochlorophyll *a* in a strain of aerobic heterotrophic bacterium. *Agric. Biol. Chem.* **42**, 1627–1628 (1987).
6. Harashima, K., Shiba, T. & Murata, N. *Aerobic Photosynthetic Bacteria*. (Springer, Berlin, 1989).
7. Shimada, K. in *Anoxygenic Photosynthetic Bacteria* (eds. Blankenship, R. E., Madigan, M. T. & Bauer, C. E.) 105–122 (Kluwer, Dordrecht, 1995).
8. Woese, C. R. et al. The phylogeny of purple bacteria; The alpha subdivision. *Syst. Appl. Microbiol.* **5**, 315–326 (1984).
9. Yurkov, V. et al. Phylogenetic positions of novel aerobic, bacteriochlorophyll *a* containing bacteria and description of *Roseococcus thiosulfatophilus* gen. nov., sp. nov., *Erythromicrobium ramosum* gen. nov., sp. nov. and *Erythrobacter litoralis* sp. nov. *Int. J. Syst. Bacteriol.* **44**, 427–434 (1994).
10. Nishimura, Y. et al. DNA relatedness and chemotaxonomic feature of aerobic bacteriochlorophyll-containing bacteria isolated from coast of Australia. *J. Gen. Appl. Microbiol.* **40**, 287–296 (1994).
11. Yurkov, V. V. & Beatty, T. Aerobic anoxygenic phototrophic bacteria. *Microbiol. Mol. Biol. Rev.* **62**, 695–724 (1998).
12. Noguchi, T., Hayashi, H., Shimada, K., Takaichi, S. & Tasumi, M. *In vivo* states and function of carotenoids in an aerobic photosynthetic bacterium, *Erythrobacter longus*. *Photosynth. Res.* **31**, 21–30 (1992).
13. Garcia, D., Mathis, P. & Vermeglio, A. Kinetics of electron transfer between the tetrahemic cytochrome and special pair in isolated reaction centers of *Roseobacter denitrificans*. *Photosynth. Res.* **55**, 331–335 (1998).
14. Schwarze, C., Carluccio, A. V., Venturulli, G. & Labahn, A. Photo-induced cyclic electron transfer involving cytochrome *bcl* complex and reaction center in the obligate aerobic phototroph *Roseobacter denitrificans*. *Eur. J. Biochem.* **267**, 422–433 (2000).
15. Shiba, T. Utilization of the light energy by the strictly aerobic bacterium *Erythrobacter* sp. OCH114. *J. Gen. Appl. Microbiol.* **30**, 1313–1320 (1984).
16. Yurkov, V. V., Krieger, S., Stackebrandt, E. & Beatty, J. T. *Citromicrobium bathymonarium*, a novel aerobic bacterium isolated from deep sea hydrothermal vent plume waters that contains photosynthetic pigment-protein complexes. *J. Bacteriol.* **181**, 4517–4525 (1999).
17. Nisbet, E. G., Cana, J. R. & Van Dover, C. L. Origins of photosynthesis. *Nature* **373**, 479–480 (1995).
18. Van Dover, C. L., Reynolds, G. T., Chave, A. D. & Tyson, J. A. Light at deep sea hydrothermal vents. *Geophys. Res. Lett.* **23**, 2049–2052 (1996).
19. Kolber, Z. S., Prasil, O. & Falkowski, P. G. Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and experimental protocols. *Biochim. Biophys. Acta* **1367**, 88–106 (1998).
20. Prasil, O., Adir, N. & Ohad, I. in *The photosystems: Structure, Function and Molecular Biology* (ed. Barber, J. R.) 295–348 (Elsevier, New York, 1992).
21. Binder, B. Cell cycle regulation and the timing of chromosome replication in marine *Synechococcus* (cyanobacteria) during light- and nitrogen-limited growth. *J. Phycol.* **36**, 120–126 (2000).
22. Behrenfeld, M. J. & Kolber, S. Z. Widespread iron limitation of phytoplankton in the South Pacific Ocean. *Science* **283**, 840–843 (1999).
23. Kolber, Z. & Falkowski, P. G. Use of active fluorescence to estimate phytoplankton photosynthesis *in situ*. *Limnol. Oceanogr.* **38**, 1646–1665 (1993).
24. Durnford, D. G. & Falkowski, P. G. Chloroplast redox regulation of nuclear gene transcription during photoacclimation. *Photosynth. Res.* **53**, 229–241 (1997).
25. Shiba, T., Shioi, Y., Takamiya, K., Sutton, D. C. & Wilkinson, C. R. Distribution and physiology of aerobic bacteria containing bacteriochlorophyll *a* on the east and west coast of Australia. *Appl. Environ. Microbiol.* **57**, 295–300 (1991).
26. Yurkov, V. & Beatty, T. Isolation of aerobic anoxygenic photosynthetic bacteria from black smoker plume waters of the Juan de Fuca Ridge in the Pacific Ocean. *Appl. Environ. Microbiol.* **64**, 337–341 (1998).
27. Wiessner, W. in *Photobiology of Microorganisms* (ed. Halldall, P.) 95–133 (Wiley-Interscience, London, 1970).
28. Droop, M. R. in *Algal Physiology and Biochemistry* (ed. Stewart, W. D. P.) 531–599 (Blackwell Scientific, Berkeley, 1974).

Acknowledgements

This research was support by the National Science Foundation and the Office of Naval Research. We thank C. Vetrani, R. Lutz, J. Henderson, G. Ananyev and D. Klimov for technical assistance and the crew of the *Alvin* for their cooperation. We thank P. Minnett for the SST data.

Correspondence and requests for materials should be addressed to P.F. (e-mail: falko@imcs.rutgers.edu).

Large-scale processes and the Asian bias in species diversity of temperate plants

Hong Qian* & Robert E. Ricklefs†

* Department of Forest Sciences, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

† Department of Biology, University of Missouri-St. Louis, St. Louis, Missouri 63121, USA

An important issue in the study of biodiversity is the extent to which global patterns of species richness reflect large-scale processes and historical contingencies^{1,2}. Ecological interactions in local assemblages may constrain the number of species that can coexist^{3,4}, but differences in diversity in similar habitats within different regions (diversity anomalies) suggest that this limit is not firm. Variation in rate of species production could influence regional and perhaps local diversity independently of the ecological capacity of an area to support coexisting species, thereby creating diversity anomalies^{5,6}. Temperate Zone genera of plants that are disjunct between similar environments in eastern Asia and eastern North America (EAS-ENA) have twice as many species in Asia as in North America⁷. Because lineages of these genera in Asia and North America are mostly sister pairs⁸, they share a common history of adaptation and ecological relationship before disjunction. Thus, the diversity anomaly in EAS-ENA genera is not an artefact of taxon or habitat sampling but reflects differences in the net diversification (speciation–extinction) of the lineages in each of the continents. Here we propose that the most probable cause of the EAS-ENA anomaly in diversity is the extreme physiographical heterogeneity of temperate eastern Asia, especially compared with eastern North America, which in conjunction with climate and sea-level change has provided abundant opportunities for evolutionary radiation through allopatric speciation.

We focus on sister pairs of disjunct congeneric lineages of temperate plants in eastern Asia and eastern North America to reduce differences in age and ecology being major factors contributing to differences in diversity. Most disjunctions between Asia and North America are thought to have formed when cooling climates in the mid and late Tertiary forced temperate-climate

plants south of the Bering Land Bridge, across which they formerly had continuous distributions^{9,10}. Of about 100-plus disjunct genera found uniquely in temperate eastern Asia and North America¹¹, 58 are restricted in North America to moist environments in the eastern half of the continent (EAS-ENA disjuncts). The ecological distributions of American and Asian counterparts are more similar in EAS-ENA genera than in disjunct genera that also occur in drier environments in western North America. Our focus on EAS-ENA disjuncts is reinforced by a TWINSPLAN partitioning of latitude–longitude grid cells based on the geographical occurrence of disjunct genera. This analysis identifies regions in eastern Asia and eastern North America (cluster III cells) to which most of the EAS-ENA disjuncts are restricted (Fig. 1).

Of the 58 EAS-ENA disjunct genera, 33 have more species in Asia and 10 have more species in North America, with 15 genera tied ($\chi^2 = 12.3$, $P < 0.001$ with ties deleted; $\chi^2 = 9.1$, $P < 0.001$ with ties split). Analysis of covariance of the $\log_{10}$ -transformed number of species per genus, with the $\log_{10}$ -transformed area (km^2) occupied by each genus included as a covariate, indicates a regional diversity bias favouring Asia by $0.30 \pm 0.08 \log_{10}$ units, which is a factor of 2.0 ($F_{1,113} = 13.5$, $P = 0.0004$). Thus, with geographical area controlled, sister lineages occupying broadly similar environments on two continents have at present twice as many species, on average, in eastern Asia as in eastern North America.

The area occupied by cluster III grid cells is physiographically more heterogeneous in eastern Asia than in eastern North America.

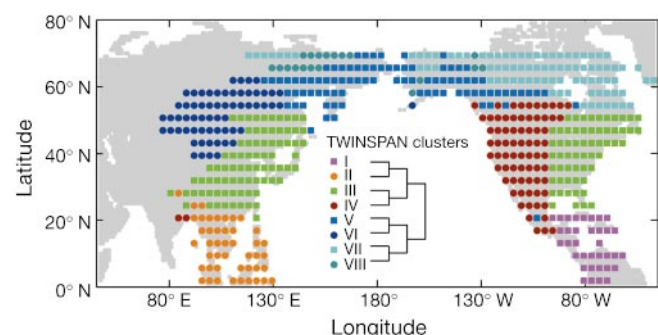


Figure 1 Geographical distribution of disjunct genera of vascular plants in eastern Asia and North America (range south of the equator not shown). Each symbol represents a grid cell with 3.75° latitude and 3.75° longitude. Symbols differentiate the eight clusters of grid cells identified by a TWINSPLAN analysis. TWINSPLAN divides grid cells in a dichotomous hierarchy based on ordination of the cells with respect to the occurrence of genera. Cluster III includes 79 grid cells in eastern Asia and 59 in eastern North America. Most of the species of EAS-ENA disjunct genera are restricted to cluster III grid cells.

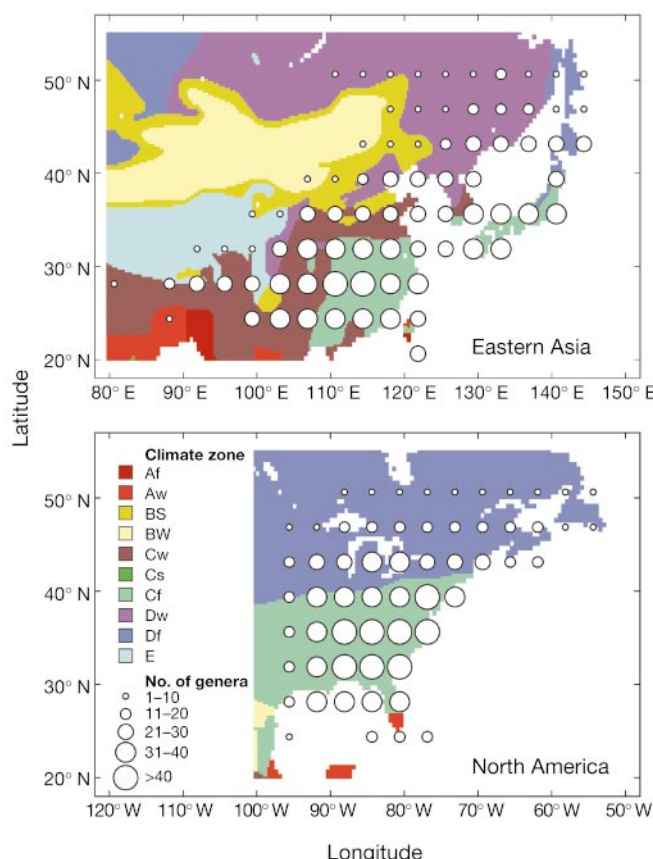


Figure 2 Maps showing climate heterogeneity and spatial patterns in the number of EAS-ENA disjunct genera of vascular plants for grid cells in TWINSPLAN cluster III. Climate classification is according to ref. 12. Climate zones: Af, tropical rainforest climate; Aw, tropical savanna climate; BS, steppe climate; BW, desert climate; Cf, warm temperate climate with all seasons moist; Cw, warm temperate rainy climate with dry winter; Cs, warm temperate rainy climate with dry summer; Dw, cold snowy forest climate with dry winter; Df, cold snowy forest climate with all seasons moist; E, ice climates.

Table 1 Distribution of genera in North America according to size classes of species richness

Number of species	1	2	3–5	6–10	11–20	21–50	51–100	>100
East only (EAS-ENA)	31	10	12		2	3		
East and west			5	3	3		1	1

The Asian region includes four Köppen climate zones¹², supporting temperate forest (warm and cool temperate climate groups C and D), distributed in 13 geographically distinct patches, including areas isolated by water on the Korean Peninsula and Japan (Fig. 2). Extensive tropical and subtropical forests at lower elevation separate temperate forests at higher elevations in southern mainland China. The comparable floristic region in North America includes only two climate zones distributed in two large adjoining areas.

The contemporary fragmentation of temperate forest in eastern Asia combined with fluctuations of climate and sea level could create conditions for generating diversity through allopatric speciation. During cooler 'glacial' climates in the Pleistocene, temperate forests would have extended to lower elevations in the south of China, as they did elsewhere in the world¹³, allowing isolated patches to join together; warmer climates would have forced temperate forest back to 'refugia' at higher elevation. Shallow seas lying between mainland China, peninsular Korea and Japan create additional isolated areas of temperate forest in Asia. During Pleistocene periods of lowered sea level associated with glacial maxima, the Yellow Sea was largely drained and these areas were connected by dry land with a climate suitable for temperate forest. CLIMAP reconstruction (<http://ingrid.ldgo.columbia.edu/SOURCES/CLIMAP/LGM/vegetation/>) of plant formations during the glacial maximum 18 thousand years ago, which is probably representative of glacial maxima throughout the Pleistocene, places deciduous forest in coastal areas between present-day China, Korea and Japan, and over much of coastal China. Eustatic fluctuations in sea level that would have alternately separated and joined plant distributions also occurred well before the Pleistocene, with pronounced sea-level highs at about 15 million and 5 million years (Myr) ago and frequent excursions between high and low levels during the past 10 Myr^{14,15}. The larger proportion of species-rich genera among those whose distributions include Korea or Japan suggests a potential role of sea-level fluctuations in species production. Of the 18 EAS-ENA genera restricted to mainland China, only 2 have more than 5 species; of 40 genera including Korea, Japan or both in their distributions, 21 have more than 5 species (likelihood-ratio $\chi^2 = 10.0$, $P = 0.002$) and 11 have more than 10 species.

Climate and geography have evidently had a much smaller role in speciation in EAS-ENA disjunct genera in the topographically more homogeneous region of eastern North America, where only 5 of 58 EAS-ENA genera have more than 5 species. Because physiographic heterogeneity is greater in western North America than in the east, however, we would expect greater species richness in genera that include western North America (TWINSPAN cluster IV) in their geographical distributions. Thirteen eastern Asia/North America herbaceous disjunct genera that occur in temperate eastern North America extend their ranges throughout the west. Of these, 8 have more than 5 species (likelihood-ratio $\chi^2 = 18.3$, $P < 0.001$; Table 1).

This difference is independent of range size. In an analysis of covariance of the $\log_{10}$ -transformed number of species per genus, with the $\log_{10}$ -transformed area (in km²) occupied by each genus included as a covariate, the species richness of genera distributed widely across North America exceed that of genera restricted to the east (EAS-ENA disjuncts) by $0.55 \pm 0.14 \log_{10}$ units, a factor of 3.5 ($F_{1,68} = 15.6$, $P = 0.0002$). Most of the diversity in these genera is concentrated in the western half of North America. The 13 genera distributed continuously from east to west across North America have 40 species restricted to the east, 23 species occurring in both the east and the west, and 298 species restricted to the west¹⁶.

The conclusion that EAS-ENA disjuncts are more diverse in Asia than in North America depends on the application of uniform taxonomic practices in the two continents. It is possible that taxonomists have split Asian disjunct species more than those in North America by applying names to poorly differentiated allopatric populations. Three observations argue against this being a factor in the Asia diversity bias among EAS-ENA disjuncts. First, disjunct genera in boreal environments (TWINSPAN clusters V–VIII) are more diverse in North America than in Asia^{17,18}. Second, greater diversity is associated with greater geographical heterogeneity in North America, as it is between eastern North America and eastern Asia. Third, much of the diversity in eastern Asia occurs within small regions of sympatry and thus is unlikely to represent artificial splitting. Using accounts in volumes published to date of the Flora of China²⁹, we examined the geographical distribution of species in four species-rich disjunct genera among the 33 provinces of China. These accounts indicate that individual species occupy 20–50% of the province-level distribution of each species in China and that up to 40–100% of the congeneric species coexist in a single province (Table 2).

The hypothesis that geographical heterogeneity combined with climate and sea-level change have had a major role in the diversification of the eastern Asian flora assumes allopatric mechanisms of species production. Plants are well known for sympatric production of new species through autopolyploidy and allopolyploidy. However, relative uniformity of chromosome numbers in genera, indicates that sympatric speciation is a minor factor in the diversification of the Asian and North American disjunct floras. A brief survey showed that chromosome numbers for 57 primarily Asian species in 14 EAS-ENA disjunct genera were the base number (2N) for 43 species, 4N for 9 species and 6N for 5 species. Thus, a minority of speciation events has involved autopolyploidy or hybridization.

Compared with the allopatric speciation hypothesis, two alternative explanations for the Asian diversity bias seem less likely. First, if Asian species were more frequently paraphyletic with respect to congeners in North America, then the Asian diversity bias could be attributed to the longer occurrence of presently disjunct genera in Asia as compared with North America. Asian paraphyly has been reported in molecular phylogenetic studies of some non EAS-ENA disjunct genera, such as *Aesculus*¹⁹ and *Gleditsia*²⁰; however, the fossil record suggests that most EAS-ENA disjuncts have longer histories in North America than in Asia. Of nine woody EAS-ENA disjunct genera²¹, all appear earlier in the fossil record of North America than in that of Asia.

Table 2 Distribution among Chinese provinces of species in four genera of species-rich EAS-ENA disjunct genera

Genus	Family	Number of species in mainland China	Number of provinces occupied by genus	Maximum number of species per province	Average number of provinces per species
<i>Meehania</i>	Lamiaceae	7	12	7	5.6
<i>Osmanthus</i>	Oleaceae	23	14	10	3.6
<i>Castanopsis</i>	Fagaceae	58	14	28	3.8
<i>Lithocarpus</i>	Fagaceae	123	17	79	3.0

Second, extinction of species associated with late-Pliocene and Pleistocene climate cooling and glaciation may have been more prevalent in North America than in eastern Asia, where extensive temperate refuges existed in subtropical latitudes during glacial maxima. However, although Pliocene-Pleistocene extinctions reduced the temperate woody flora of Europe considerably^{22,23}, extinctions were less important in eastern North America⁶. Nonetheless, greater extinction of genera in North America (18% of moist-temperate tree genera represented in the mid-Tertiary fossil record) than in Asia (4%) may translate into more pronounced extinction of species in extant genera. Thus, North American extinctions may have had a role in the diversity bias favouring eastern Asia. If so, however, they did not reverse the diversity bias favouring North America in disjunct genera with more northerly and westerly distributions in North America. Moreover, a twofold Asian diversity bias was already present with respect to genera of woody plants (122 versus 60 in eastern North America⁶) before late Tertiary climate cooling.

The role of alternate fragmentation and rejoining of geographical ranges in species production has been discussed in the past, most famously in regard to adaptive radiation in island archipelagos where fragmentation and secondary sympatry result from infrequent migration between islands. The temperate disjunct flora of eastern Asia is a superior continental model for increase in regional diversity through allopatric speciation, for several reasons. First, habitat fragmentation and rejoining are based on observed effects of climate change on elevational distribution of temperate environments^{13,24} and on documented changes in sea level¹⁴. Second, comparisons of diversity between physiographically complex (eastern Asia) and simple (eastern North America) regions are based on sister taxa, thereby controlling for age and, to a large extent, for ecological relationships of plants to their physical environments. Third, the diversity anomaly between EAS-ENA disjunct genera occurs at the level of species in genera and is likely to have been produced by means of allopatric speciation under the influence of climate cycles associated with general climate cooling and glaciation during the late Tertiary. Distributions of temperate disjunct taxa in the fossil record suggest that temperate connections between Asia and North America were severed by the mid to late Miocene (5–10 Myr ago)^{25,26}. Genetic divergence has been coupled with variously calibrated molecular clocks⁸ to place ages of disjunctions between 2 and 29 Myr, with half the estimates falling between 4 and 12 Myr, providing ample time for independent diversification of lineages in each region.

The species richness of disjunct EAS-ENA genera is plausibly related to opportunities for repeated cycles of allopatry and secondary sympatry. Particularly in eastern Asia, these cycles have been driven by changes in climate and sea level within a physiographically heterogeneous region. High levels of species richness in genera are associated with distribution across currently fragmented geographical areas in eastern Asia. In eastern North America, more uniform climate and simpler geography have not fostered evolutionary radiation among the same lineages as in eastern Asia. Only among disjunct herbaceous genera that include drier environments of western North America in their habitat distribution has speciation proceeded at a rapid rate. We suggest that history and large-scale processes may explain much of the variation in species richness over the globe. □

Methods

Plant distributions

We reviewed published lists of eastern Asian/North American disjunct genera of vascular plants and included 102 genera in our analysis. Of these, 58 genera are restricted to eastern North America. Some of these extend their distributions to tropical areas in Asia and America. To document generic diversity on a smaller scale, we divided the region between 45° E and 45° W longitude and between 15° S and 75° N latitude into 1728 grid cells 3.75° on a side. Presence/absence of each disjunct genus in each grid cell was documented

according to distribution maps^{11,27} and additional floristic data. Within the grid area, 580 cells, 289 in eastern Asia and 291 in North America, had at least one of the 102 disjunct genera. We divided the 580 grid cells into clusters using a two-way indicator species, hierarchical, divisive cluster approach (TWINSPAN²⁸). Three hierarchical levels of clustering were performed, which resulted in eight clusters (I–VIII).

Chromosome numbers

A sample of chromosome numbers for species in disjunct genera was obtained from volumes of the Index to Plant Chromosome Numbers³⁰. Chromosome numbers were included only when two or more species were reported for a continent. Most of the species in the sample were eastern Asian.

Received 25 January; accepted 14 July 2000.

- Ricklefs, R. E. Community diversity: relative roles of local and regional processes. *Science* **235**, 167–171 (1987).
- Ricklefs, R. E. & Schluter, D. in *Species Diversity in Ecological Communities* (eds Ricklefs, R. E. & Schluter, D.) 350–363 (Univ. Chicago Press, Chicago, 1993).
- MacArthur, R. H. *Geographical Ecology: Patterns in the Distribution of Species* (Harper and Row, New York, 1972).
- Currie, D. J. Energy and large-scale patterns of animal species and plant species richness. *Am. Nat.* **137**, 27–49 (1991).
- Ricklefs, R. E. & Latham, R. E. in *Species Diversity in Ecological Communities* (eds Ricklefs, R. E. & Schluter, D.) 215–229 (Univ. Chicago Press, Chicago, 1993).
- Latham, R. E. & Ricklefs, R. E. in *Species Diversity in Ecological Communities* (eds Ricklefs, R. E. & Schluter, D.) 294–314 (Univ. Chicago Press, Chicago, 1993).
- Qian, H. & Ricklefs, R. E. A comparison of the taxonomic richness of vascular plants in China and the United States. *Am. Nat.* **154**, 160–181 (1999).
- Wen, J. Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. *Annu. Rev. Ecol. Syst.* **30**, 421–455 (1999).
- Graham, A. *Late Cretaceous and Cenozoic History of North American Vegetation North of Mexico* (Oxford Univ. Press, New York, 1999).
- Tiffney, B. H. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. *J. Arnold Arbor.* **66**, 73–94 (1985).
- Hong, D.-Y. Eastern Asian–North American disjunctions and their biological significance. *Cathaya* **5**, 1–39 (1993).
- Müller, M. J. *Selected Climate Data for a Global Set of Standard Stations for Vegetation Science* (Dr. W. Junk Publishers, The Hague, 1982).
- Bush, M. B., Weimann, M., Piperno, D. R., Liu, K.-B. & Colinvaux, P. A. Pleistocene temperature change and vegetation depression in Ecuadorian Amazonia. *Quat. Res.* **34**, 330–345 (1990).
- Haq, B. U., Hardenbol, J. & Vail, P. R. Chronology of fluctuating sea levels since the Triassic. *Science* **235**, 1156–1167 (1987).
- Kominz, M. A., Miller, K. G. & Browning, J. V. Long-term and short-term global Cenozoic sea-level estimates. *Geology* **26**, 311–314 (1998).
- Kartesz, J. T. A *Synonymized Checklist of the Vascular Flora of the United States, Canada, and Greenland* (Timber, Portland, 1994).
- Li, S.-Y. & Adair, K. T. *Species Pools of Seed Plants in Eastern Asia and North America* (Arthur Temple College of Forestry, Stephen F. Austin State Univ., Nacogdoches, TX, 1997).
- Qian, H. Spatial pattern of vascular plant diversity in North America north of Mexico and its floristic relationship with Eurasia. *Ann. Bot.* **83**, 271–283 (1999).
- Xiang, Q.-Y., Crawford, D. J., Wolfe, A. D., Tang, Y.-C. & DePamphilis, C. W. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): a molecular phylogenetic perspective. *Evolution* **52**, 988–997 (1998).
- Schnabel, A. & Wendel, J. E. Cladistic biogeography of *Gleditsia* (Leguminosae) based on *ndhF* and *rpl16* chloroplast gene sequences. *Am. J. Bot.* **85**, 1753–1765 (1998).
- Manchester, S. R. Biogeographical relationships of North American Tertiary floras. *Ann. Missouri Bot. Gard.* **86**, 472–522 (1999).
- Sauer, J. D. *Plant Migration: The Dynamics of Geographic Patterning in Seed Plant Species* (Univ. California Press, Berkeley, 1988).
- Mai, D. H. *Tertiäre Vegetationsgeschichte Europas* (G. Fischer, Jena, Germany, 1995).
- Hooghiemstra, H. Quaternary and upper-Pliocene glaciations and forest development in the tropical Andes: evidence from a long-high-resolution pollen record from the sedimentary basin of Bogotá, Colombia. *Palaeogeogr. Palaeoclim. Palaeoecol.* **72**, 11–26 (1989).
- Tiffney, B. H. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *J. Arnold Arbor.* **66**, 243–273 (1985).
- Wolfe, J. A. Some aspects of plant geography of the Northern Hemisphere during the Late Cretaceous and Tertiary. *Ann. Missouri Bot. Gard.* **62**, 264–279 (1975).
- Li, H.-L. Floristic relationships between eastern Asia and eastern North America. *Trans. Am. Philos. Soc. New Series* **42**, 371–429 (1952).
- Hill, M. O. TWINSPAN: a FORTRAN Program for Arranging Multivariate Data in a Two-way Table by Classification of Individuals and Attributes (Cornell Univ., Ithaca, New York, 1979).
- Wu, Z.-Y. & Raven, P. H. (eds) *Flora of China* Vol. 4, Vol. 15 & Vol. 17, (Missouri Botanical Garden Press, St. Louis; Science Press, Beijing, 1994, 1996 & 1999).
- Goldblatt, P. (ed) *Index to Plant Chromosome Numbers 1975–1978, 1979–1981, 1982–1983 Monographs in Systematic Botany* Vol. 5, Vol. 8 & Vol. 13, (Missouri Botanical Garden Press, 1981, 1984 & 1985).

Acknowledgements

We thank E. Bermingham, Q.-F. Guo, S. Heard, J. Losos, I. Lovette, S. Manchester, S. Renner, D. Schluter and two anonymous reviewers for valuable comments on the manuscript.

Correspondence and requests for materials should be addressed to R.E.R. (e-mail: ricklefs@umsl.edu).

Genetic evidence for female host-specific races of the common cuckoo

H. Lisle Gibbs*, Michael D. Sorenson†, Karen Marchetti*‡,
M. de L. Brooke§, N. B. Davies§ & Hiroshi Nakamura||

* Department of Biology, McMaster University, Hamilton, Ontario L8S 4K1, Canada

† Department of Biology, Boston University, Boston, Massachusetts 02215, USA

§ Department of Zoology, Downing Street, University of Cambridge, Cambridge CB2 3EJ, UK

|| Faculty of Education, Shinshu University, Nagano 380, Japan

The common cuckoo *Cuculus canorus* is divided into host-specific races (gentes)¹. Females of each race lay a distinctive egg type that tends to match the host's eggs, for instance, brown and spotted for meadow pipit hosts or plain blue for redstart hosts^{2–4}. The puzzle is how these gentes remain distinct. Here, we provide genetic evidence that gentes are restricted to female lineages, with cross mating by males maintaining the common cuckoo genetically as one species. We show that there is differentiation between gentes in maternally inherited mitochondrial DNA, but not in microsatellite loci of nuclear DNA. This supports recent behavioural evidence that female, but not male, common cuckoos specialize on a particular host⁵, and is consistent with the possibility that genes affecting cuckoo egg type are located on the female-specific W sex chromosome⁶. Our results also support the ideas that common cuckoos often switched hosts during evolution^{7,8}, and that some gentes may have multiple, independent origins, due to colonization by separate ancestral lineages.

We assessed patterns of variation in cuckoo populations in Great Britain⁹ and Japan⁵ using two types of rapidly-evolving genetic markers with different modes of inheritance. If gentes of the common cuckoo are restricted to female lineages, then we would expect significant differences between their mitochondrial DNA (mtDNA), as it is transmitted through females only. However, their nuclear microsatellite DNA (nDNA), which segregates through both sexes, would not differ. On the other hand, if male cuckoos mate only with female cuckoos raised by the same host, then the gentes would be genetically isolated (cryptic species) and would differ in both mtDNA and nDNA.

We obtained blood samples from unrelated cuckoo chicks found in nests across Great Britain with three major hosts—the reed warbler (*Acrocephalus scirpaceus*), the meadow pipit (*Anthus pratensis*) and the dunnoek (*Prunella modularis*). Average egg colour patterns differ among these gentes³. We were unable to document the colour of the eggs from which the chicks hatched and so our analyses refer to cuckoo chicks raised in different host nests and not host-specific egg patterns. We assume that each chick hatched from an egg whose morphology is representative of a particular cuckoo race.

Figure 1a shows an unrooted, maximum parsimony tree of relationships between the 20 mtDNA control region haplotypes (Table 1) from 43 cuckoo chicks in Great Britain. With one exception (GB 5), individual haplotypes are found in only a single type of host (Table 1). Haplotypes within a given gens can be quite different (for example, 'reed warbler' cuckoo haplotypes GB1 and GB3) leading to a pattern in which host-specific haplotypes form multiple clusters of closely-related haplotypes separated by a number of mutational steps (Fig. 1a). The frequencies of different haplotypes are significantly different between cuckoo chicks from different hosts (exact test across all three hosts, $P < 0.001$; all pair-

wise exact tests, $P < 0.05$) and values were all significantly different from zero (overall $F_{ST} = 0.18$; $P < 0.001$; pair-wise F_{ST} values = 0.10–0.27; all $P < 0.05$) (see analysis section in methods).

The genetic similarity in haplotypes among chicks from the same host race cannot be explained as a by-product of sampling large numbers of related individuals from the same locality. Four of the seven haplotypes with more than one chick (GB 2, GB 5, GB 8 and GB 17) were represented by samples collected from geographically distinct locations separated by a mean ($\pm$ s.d.) of 180.8 ± 86.5 km (range, 10–324 km). The three others (GB 1, GB 3 and GB 15) had chicks from more than one site (9, 2 and 2 sites respectively; mean distance apart, 147.8 ± 66.7 km; range, 10–280 km) and small numbers of chicks (≤ 5) from the same site. Multiple chicks from the same site were unrelated on the basis of genetic and spatial criteria⁹. In addition, when analyses were conducted for samples from each race separately to control for the effect of host type, there was no relationship between pair-wise genetic distances between individual sequences and geographical distances between sample locations (Mantel tests, all $P > 0.05$; range of r^2 values, 0.008–0.08).

Japanese cuckoos from a single population near Nagano^{5,10} represent an interesting comparison with the British cuckoos in two ways. First, although there is behavioural evidence from genetic analysis of parentage⁵ and radio-tracking¹⁰ that females are host specialists, individuals with different hosts lay eggs that are more similar and mimic host eggs less well than those in Great Britain¹¹. Second, one of the main hosts of this population, the azure-winged magpie (*Cyanopica cyana*), only became a host about 30 years ago¹¹.

Figure 1b shows an unrooted, maximum parsimony tree of the relationships between the four mtDNA control region haplotypes

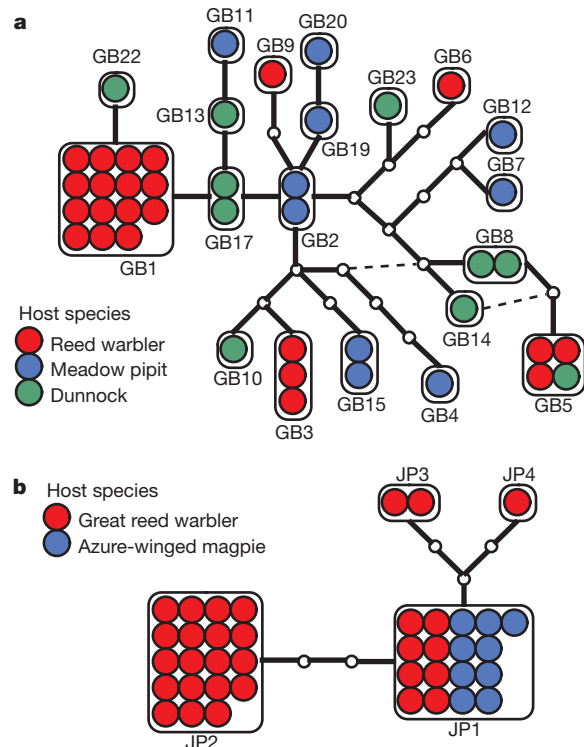


Figure 1 Unrooted, maximum parsimony trees for cuckoo mtDNA haplotypes from Great Britain (a) and Japan (b). Individual haplotypes are represented by boxes labelled with the name of the haplotype (Great Britain, GB1 to GB23; Japan, JP1 to JP4). The number of coloured circles within each box shows the number and host identity of cuckoos with a particular haplotype. The number of line segments connecting the boxes gives the number of nucleotide differences between two haplotypes. For the GB network, alternative connections defining other equally parsimonious trees are shown by dotted lines.

‡ Present address: Department of Biology 0116, University of California at San Diego, La Jolla, California 92093, USA.

found in the 39 female cuckoos of the Japanese population. Two common (JP1 and 2) and two rare (JP3 and 4) haplotypes were present (Table 1). Mean sequence divergence between these haplotypes and those from Great Britain equals 5.5% (range, 4.6–6.3%). The reduced level of mtDNA variation in the Japanese cuckoos may be due to the restricted area from which they were collected as compared with the British birds. As with the British cuckoos, there was a highly significant difference ($P < 0.0001$) in haplotype frequencies of cuckoos from different races; all ‘magpie’ females had the common haplotype JP1, whereas the ‘great reed warbler, (*Acrocephalus arundinaceus*)’ females possessed all four haplotypes (Table 1). The pair-wise F_{ST} value was also large and significant ($F_{ST} = 0.39$; $P < 0.001$). This differentiation is not simply due to the presence of related groups of females with the same host. After adjusting for the presence of possible relatives among the sampled birds, a total of 23 ‘great reed warbler’ and eight ‘magpie’ cuckoos remained. Nonetheless, the pair-wise F_{ST} value remained large and significant ($F_{ST} = 0.30$; $P = 0.0098$) as did the differences in haplotype frequencies ($P = 0.0037$). The limited variation in the ‘magpie’ cuckoos is consistent with the idea that, although such cuckoos are behaviourally a distinct race, they are recently derived. However, this pattern could also be due to the smaller number of ‘magpie’ cuckoos sampled. Their distinctiveness, despite their recent origin, is expected if the ‘magpie’ race in Nagano is composed of colonists from other pre-existing gentes in the area¹¹ or is made up of females from extant ‘magpie’ races in other parts of Japan.

Assuming that identical mitochondrial sequences did not originate independently in different gentes by parallel (or convergent) mutations, there has been a minimum of nine host switches by British female cuckoos. Including the host as an equally weighted character lowers this to eight host switches. A single host switch from great reed warblers to azure-winged magpies is consistent with the available mtDNA data for Japanese cuckoos. In both populations, the minimum number of host switches inferred from the mtDNA haplotype tree is far less than that expected under a null model of no association between cuckoo mtDNA lineages and host species ($P \ll 0.001$).

In contrast to the mtDNA results, no significant differences were observed in allele frequencies (all $P > 0.12$ for exact tests) nor was the overall value (overall value 0.0; $P = 0.53$; range of locus-specific R_{ST} values, 0.0–0.05) significantly different from zero for any of eight nuclear microsatellite loci between ‘magpie’ or ‘great reed warbler’ cuckoos in Japan (Table 2). For the British cuckoo chicks, only 1 of 30 host-race-by-locus exact tests (cuckoo-specific microsatellite locus Ccp.) 02 for the dunnoek versus reed warbler comparison) was significant ($P < 0.01$) after adjusting the critical P value for judging significance using a sequential Bonferroni procedure. Host-race-by-locus R_{ST} values ranged from 0.0 to 0.19, though the overall R_{ST} value (0.023; $P = 0.12$) (Table 2) and all pair-wise values were small and not significantly different from zero (dunnoek versus meadow pipit, $R_{ST} = 0.0012$; $P = 0.53$; dunnoek versus reed warbler, $R_{ST} = 0.021$; $P = 0.22$; meadow pipit versus reed warbler, $R_{ST} = 0.038$; $P = 0.07$). Using an analysis of molecular variance (AMOVA) to compare directly the results for each class of marker, both populations show a highly significant variation in mtDNA between cuckoos associated with different hosts (GB 19% and Japan 40% of the total variation; both $P \leq 0.002$), but no significant differentiation in nuclear microsatellite loci (0% total variation for both; $P \geq 0.73$). These sex-dependent patterns of genetic differentiation demonstrate the presence of female-specific gentes in cuckoos.

Sequence divergence between mtDNA haplotypes within Great Britain and Japan was low, indicating that these gentes are young in age. The divergence in sequence between the British haplotypes ranges from 2.2% to 0.24% (mean, 1.26%). For Japan, the range is 1.46–0.73% (mean, 1.02%) (Table 1). To estimate time to the most recent common ancestor for all haplotypes for each region, we calculated the average sequence divergence and its standard deviation across the basal node for each region (that is, Great Britain and Japan) in a rooted haplotype tree¹². Haplotypes from the two regions were used as reciprocal outgroups in this analysis. Mean (± 2 s.d.) sequence divergence across the basal node for each region in the haplotype tree was 1.6% ($\pm 1.0\%$) for Great Britain and 1.3% ($\pm 0.9\%$) for Japan. Assuming a molecular clock calibration of 20%

Table 1 mtDNA haplotype variation and distribution among hosts

[illegible]

Positions of variable sites refer to nucleotide positions relative to the beginning of the 411-bp mtDNA sequence. Dots represent identical bases to the first sequence. For GB haplotypes, Hosts 1, 2 and 3 refer to reed warbler, meadow pipit and duncock hosts respectively, while for JP haplotypes, Hosts 1 and 2 refer to great reed warbler and magpie hosts, respectively. These sequences have been deposited in GenBank under accession nos. AF282694–AF282717.

divergence per Myr for this portion of the control region¹³, the most recent common ancestor of British haplotypes occurred 80,000 ($\pm 50,000$) yr ago and the most recent common ancestor of Japanese haplotypes occurred 65,000 ($\pm 45,000$) yr ago. Although these estimates are subject to error associated with assumptions about mutation rate, the time of the most recent mitochondrial ancestor provides an upper limit to the age of extant gentes.

These genetic results confirm inferences from studies of behaviour⁵ and egg morphology^{3,4} that cuckoo females, but not males, specialize on particular hosts, possibly through some type of imprinting mechanism^{14,15}, and such female-specific behaviour has continued long enough to result in an unusual pattern of sex-specific genetic differentiation in this parasitic species. Our previous attempt to test for mitochondrial differentiation between cuckoo gentes⁹ probably failed because the genetic marker used (an mtDNA control region microsatellite) has an extremely high mutation rate. Phylogenetic reconstruction of repeat number in this microsatellite in the context of the new mtDNA sequence data reveals that this marker retains little historical information.

Other studies have linked egg colour variation to loci on autosomes¹⁶ and the Z sex chromosome¹⁷. However, our results agree with the proposal that genetic factors controlling the expression of egg colour are present on the female-specific W chromosome⁶, as observed for egg patterning in the great tit *Parus major* (A. Gosler, personal communication). This assumes that the patterns of differentiation in cuckoo W chromosomes parallel those observed in mtDNA in the present study. Because we lack direct evidence on the heritability of egg colour from parent–offspring comparisons, we cannot exclude explanations based on autosomal inheritance of cuckoo egg characteristics⁵. If inheritance of these characteristics is female-dependent, then these results would add to the view that sex-determining chromosomes, like W in birds, contain loci that code for products involved in reproductive function, such as the SRY gene in mammals (for review see ref. 18). This female-specific pattern of differentiation also indicates that other possible host-specific adaptations such as matching of host begging calls¹⁹ must have a significant environmental (for example, learned) component as they would be expressed in both male and female cuckoo chicks.

The mtDNA data indicate a rate of host-switching that is very low over ecological time, on the scale of years and decades, such that cuckoo populations associated with different hosts develop significant differences in mtDNA haplotype frequencies. Over longer time periods, however, this low rate of host switching prevents the differentiation of cuckoo gentes into distinct mtDNA clades. Our results support the ideas that gentes are, in an evolutionary sense, relatively young, that host switching has occurred frequently^{7,11} and that, at least in Great Britain, each ‘host-specific race’ may have multiple, independent origins due to colonizations by separate ancestral lineages at different times^{8,20}. □

Table 2 Locus-specific and overall R_{ST} values for nuclear microsatellite DNA variation

Locus	R_{ST} values	
	Great Britain	Japan
Ccμ 02	0.0	0.0
Ccμ 13	0.13	0.06
Ccμ 60	0.0	0.0
Ccμ 88	0.04	0.05
Ccμ 100	0.05	0.0
Ccμ 108	0.01	0.0
Ccμ 119	0.01	0.0
Ccμ 137	0.0	0.0
Overall	0.02	0.0
P value	(0.12)	(0.53)

R_{ST} values are for comparisons among cuckoos across all hosts for each population. The P value is the probability that the overall R_{ST} value for a particular population is different from zero.

Methods

Cuckoo samples

DNA samples for this study came from previously studied cuckoo populations in Great Britain⁹ and Japan⁵. British samples were collected as described⁹ from 43 unrelated cuckoo chicks found in the nests of three major hosts: reed warbler ($n = 23$), meadow pipit ($n = 10$) and dunnoek ($n = 10$). For Japan, DNA samples were analysed from 39 adult female cuckoos from a single marked population in Nagano whose laying histories had been determined using DNA-based parentage analyses as described⁵. As these earlier results indicate that females in this population are largely host specific, we assumed that females identified as laying only a single egg in a particular host were specialists on that host. Overall 30 females were classified as specializing on great reed warblers and nine females were classified as azure-winged magpie specialists. These samples include a female who laid two eggs in magpie nests and one egg in a warbler nest (classified as a ‘magpie’ specialist) and exclude a female identified as laying one egg in the nest of each host.

Genetic data

For mtDNA analyses, we determined sequence variation in a 411-base-pair portion of the left-hand hypervariable region of the mtDNA control region using two *Cuculus*-specific primers (CCRLIA, 5'-CAT GAT ACA TTA CAT GTA TGC CTG-3' and CCRH1, 5'-CTG AAA TAG TAT GAA TGT ATC TGT G-3'). To assess nuclear microsatellite variation in the Japanese samples, we analysed unpublished data on allelic variation in the eight cuckoo-specific loci (Ccμ 02, 13, 60, 88, 100, 108, 119 and 137)²¹ which had been used for the parentage analyses⁵. For the British samples, we genotyped the samples using the cuckoo-specific loci described above (except for Ccμ 13) and combined these data with genotypes for the three previously presented loci (Ccμ 01, 09, and 13)⁹ for an overall data set for the British samples based on 10 loci.

Analyses

Phylogenetic relationships among unique mtDNA haplotypes were inferred using maximum parsimony as implemented in PAUP²². We used ARLEQUIN²³ to conduct pair-wise comparisons of mtDNA haplotype and microsatellite allele frequencies between putative gentes. Differentiation was assessed using exact tests²⁴ for differences in haplotype or allele frequencies and in terms of F_{ST} ²⁵ values for mtDNA and R_{ST} ²⁶ values for microsatellite variation. For F_{ST} values, similarity between haplotypes was weighted on the basis of the magnitude of gamma-corrected ($\alpha = 0.08$) Kimura 2-parameter values between haplotypes. We also used this program for AMOVA²⁷ to partition total variance in both mtDNA and microsatellite variation into within versus between putative host race components. Significance levels for particular tests or values were assessed using permutation procedures implemented in the respective programs.

To identify possible relatives among ‘magpie’ and ‘great reed warbler’ females from Japan, we used the microsatellite allele frequencies for the overall population²¹ with the program KINSHIP²⁸ to calculate 95% confidence intervals for the expected r values²⁹ for pair-wise comparisons of unrelated individuals. Any pair of females with the same mtDNA haplotype and an r value greater than 0.33 (the upper bound of the interval) were identified (1/30 and 8/420 possible pairs of ‘magpie’ and ‘great reed warbler’ females, respectively) and a single randomly chosen member of pair was deleted from the data set. We then repeated our analysis of differentiation in mtDNA using this modified data set.

To estimate the minimum or most parsimonious number of host switches that could have occurred given the evolutionary history of these populations, the number of steps in the character ‘host’ on the shortest mtDNA trees was determined using MacClade³⁰ with the assumption of soft polytomies. To evaluate if the observed number of host switches was smaller than expected under a null model of no association between ‘host’ and cuckoo mtDNA haplotype, 1,000 randomized characters were generated by reassigning individual cuckoos to hosts and calculating the minimum number of host switches for the randomized characters on the shortest mtDNA trees.

Received 6 January; accepted 20 June 2000.

- Newton, A. A. *Dictionary of Birds*. (Black, London, 1893).
- Baker, E. C. S. *Cuckoo Problems*. (Witherby, London, 1942).
- Brooke, M. de L. & Davies, N. B. Egg mimicry by cuckoos, *Cuculus canorus*, in relation to discrimination by hosts. *Nature* **335**, 630–632 (1988).
- Moksnes, A. & Roskaft, E. Egg-morphs and host preference in the common cuckoo (*Cuculus canorus*): an analysis of cuckoo and host eggs from European museum collections. *J. Zool.* **236**, 625–648 (1995).
- Marchetti, K., Nakamura, H. & Gibbs, H. L. Host race formation in the Common Cuckoo. *Science* **282**, 471–472 (1998).
- Punnett, R. C. Inheritance of egg-colour in the parasitic cuckoos. *Nature* **132**, 892 (1933).
- Davies, N. B. & Brooke, M. de L. An experimental study of co-evolution between the cuckoo, *Cuculus canorus*, and its hosts. II. Host egg markings, chick discrimination, and general discussion. *J. Anim. Ecol.* **58**, 225–236 (1989).
- Davies, N. B. *Cuckoos, Cowbirds and Other Cheats* (Poyser, London, 2000).
- Gibbs, H. L., Brooke, M. de L., & Davies, N. B. Analysis of genetic differentiation of host races of the common cuckoo *Cuculus canorus* using mitochondrial and microsatellite DNA variation. *Proc. R. Soc. Lond. B* **263**, 89–96 (1996).
- Nakamura, H. & Miyazawa, Y. Movements, space use and social organization of radio-tracked cuckoos during the breeding season in Japan. *Jap. J. Ornithol.* **46**, 23–53 (1997).
- Nakamura, H., Kubota, S. & Suzuki, R. In *Parasitic Birds and Their Hosts* (eds Rothstein, S. I. & Robinson, S. K.) 94–112 (Oxford Univ. Press, Oxford, 1998).
- Steel, M. A., Cooper, A. C. & Penny, D. Confidence intervals for the divergence time of two clades. *System. Biol.* **45**, 127–134 (1996).
- Avise, J. C. & Walker, D. Pleistocene phylogeographic effects on avian populations and the speciation process. *Proc. R. Soc. Lond. B* **265**, 457–463 (1998).

14. Teuschl, Y., Taborsky, B. & Taborsky, M. How do cuckoos find their hosts? The role of habitat imprinting. *Anim. Behav.* **56**, 1425–1433 (1998).
15. Brooke, M. de L. & Davies, N. B. A failure to demonstrate host imprinting in the common cuckoo *Cuculus canorus* and alternative hypotheses for the maintenance of egg mimicry. *Ethology* **89**, 154–166 (1991).
16. Collias, E. C. Inheritance of egg-colour polymorphism in the village weaver (*Ploceus cucullatus*). *Auk* **110**, 683–693 (1993).
17. Schoffier, R. N. et al. The effect of a protoporphyrin mutant on some economic traits of the chicken. *Poultry Sci.* **61**, 817–820 (1982).
18. Roldan, E. R. S. & Gomendio, M. The Y chromosome as a battleground for sexual selection. *Trends Ecol. Evol.* **14**, 58–62 (1999).
19. Kilner, R. M., Noble, D. G. & Davies, N. B. Signals of need in parent-offspring communication and their exploitation by the common cuckoo. *Nature* **397**, 667–672 (1999).
20. Dawkins, R. *Unweaving the Rainbow*. (Allen Lane, London, 1998).
21. Gibbs, H. L., De Sousa, L., Marchetti, K. & Nakamura, H. Isolation and characterization of microsatellite DNA loci for an obligate brood parasitic bird, the common cuckoo (*Cuculus canorus*). *Mol. Ecol.* **7**, 1437–1439 (1998).
22. Swofford, D. PAUP, Ver. 4.0 (Sinauer Associates, 1999).
23. Schneider, S., Kueffer, J. M., Roessli, D. & Excoffier, L. *Arlequin ver. 1.1: A Software for Population Genetic Data Analysis*. (Genetics and Biometry Laboratory, Univ. Geneva, Switzerland, 1997).
24. Raymond, M. & Rousset, F. An exact test for population differentiation. *Evolution* **49**, 1280–1283 (1995).
25. Hudson, R., Slatkin, M. & Maddison, W. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589 (1992).
26. Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462 (1995).
27. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA. *Genetics* **131**, 479–491 (1992).
28. Goodnight, K., Queller, D. C. & Poznansky, T. Kinship Ver. 1.2 (1997). Available on <http://www.bioc.rice.edu/~kfg/GSoft.html>.
29. Queller, D. C. & Goodnight, K. F. Estimating relatedness using genetic markers. *Evolution* **42**, 258–275 (1989).
30. Maddison, W. P. & Maddison, D. R. *MacClade: Analysis of Phylogeny and Character Evolution*. Ver. 3.0. (Sinauer Associates, 1992).

Acknowledgements

We thank L. De Sousa, L. M. Tabak and M. Ramon for help in the lab, and A. Baker, P. Fu, B. Golding, A. Lotem, R. A. Morton, R. B. Payne, T. D. Price and E. Røskoft for discussion and comments. Funded by grants from NSERC (Canada) (H.L.G.), Boston University (M.D.S.), SERC (UK) and NERC (UK) (N.B.D.), and a NATO Postdoctoral Fellowship (K.M.) H.L.G. thanks S. I. Rothstein for initiating part of this collaboration by introducing him to H.N.

Correspondence and requests for materials should be addressed to H.L.G. (e-mail: gibbs@mcmaster.ca).

WNT signalling molecules act in axis formation in the diploblastic metazoan *Hydra*

Bert Hobmayer*, Fabian Rentzsch*, Kerstin Kuhn*, Christoph M. Happel*, Christoph Cramer von Laue*, Petra Snyder*, Ute Rothbächer† & Thomas W. Holstein*

* Department of Molecular Cell Biology, Zoological Institute, Darmstadt University of Technology, Schnittspahnstrasse 10, 64287 Darmstadt, Germany
† Laboratoire de Genetique et Physiologie du Developpement, IBDM, Campus de Luminy, 13288 Marseille Cedex 9, France

Members of the Wnt/wingless family of secreted proteins act as short-range inducers and long-range organizers during axis formation, organogenesis and tumorigenesis in many developing tissues¹. Wnt signalling pathways are conserved in nematodes, insects and vertebrates². Despite its developmental significance, the evolutionary origin of Wnt signalling is unclear. Here we describe the molecular characterization of members of the Wnt signalling pathway—Wnt, Dishevelled, GSK3, β -Catenin and Tcf/Lef—in *Hydra*, a member of the evolutionarily old metazoan phylum Cnidaria. Wnt and Tcf are expressed in the putative *Hydra* head organizer, the upper part of the hypostome. Wnt, β -Catenin and Tcf are transcriptionally upregulated when head organizers are established early in bud formation and head

regeneration. Wnt and Tcf expression domains also define head organizers created by *de novo* pattern formation in aggregates. Our results indicate that Wnt signalling may be involved in axis formation in *Hydra* and support the idea that it was central in the evolution of axial differentiation in early multicellular animals.

We used polymerase chain reaction (PCR) to identify and clone Wnt, *Dishevelled* and GSK3 orthologues from the fresh water polyp *Hydra* (*HyWnt*, *HyDsh* and *HyGSK3*, respectively). *Hydra Tcf/Lef* (*HyTcf*) was isolated in a yeast two-hybrid screen with a *Hydra* β -Catenin³ (*Hy β -Cat*) bait. A *Hydra* member of the Frizzled family of transmembrane receptors has also been identified⁴. *HyWnt* has a relatively low overall amino-acid identity (about 35%) to other Wnt factors, but exhibits a Wnt-specific pattern of 22 cysteines and a signal peptide sequence indicating that it is secreted (Fig. 1a). The cytoplasmic members of the pathway are highly conserved: *HyDsh* has DIX, PDZ and DEP domains, *HyGSK3* has a conserved catalytic domain, and *Hy β -Cat* has 13 armadillo repeats and a presumptive GSK3 phosphorylation site (Fig. 1a and ref. 3). The transcription factor *HyTcf* contains a high mobility group (HMG) box most closely related to HMG boxes in vertebrate Tcf proteins and a putative β -Catenin binding domain at its amino-terminal end (Fig. 1a). Predicted amino-acid sequence alignments of *HyWnt*, *HyDsh*, *HyGSK3* and *HyTcf* are presented in the Supplementary Information.

The high degree of structural conservation in protein–protein interaction domains in the cytoplasmic transducers indicates that a Wnt signalling pathway evolved early in metazoan evolution. To provide evidence for functional conservation, we examined the ability of *Hy β -Cat* to induce secondary body axes in *Xenopus* embryos. Ectopic expression of *Hy β -Cat* by injection of messenger RNA into ventral blastomeres at the eight-cell stage induced 100% ($n = 20$) complete secondary body axes containing the most anterior structures (such as eyes and cement gland), indistinguishable from those induced by *Xenopus* β -Catenin (Fig. 1b). *Hy β -Cat* can therefore interact with endogenous transcription factors, presumably

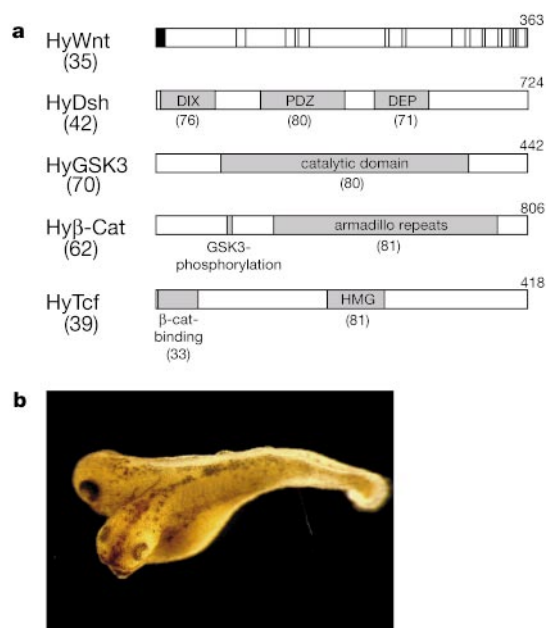


Figure 1 Structural and functional conservation of *Hydra* Wnt signalling molecules. **a**, Schematic representation of the predicted *HyWnt*, *HyDsh*, *HyGSK3*, *Hy β -Cat* and *HyTcf* proteins based on cDNA sequences. *HyWnt* has a signal peptide sequence (black box) and 22 conserved cysteine residues (vertical lines). Numbers at the ends of the bars represent the number of amino acids in the corresponding open reading frame (ORF). Numbers in parentheses show amino-acid identity between the *Hydra* and human orthologues (*Wnt1*, *Dishevelled1*, *GSK3 β* , β -Catenin and *Tcf4*) throughout the ORF and in protein–protein interaction domains. For details in the predicted *Hy β -Cat* structure see ref. 3. **b**, *Hy β -Cat* induces complete secondary axes in *Xenopus* embryos.

XTcf-3 (ref. 5), to regulate transcription of axis-inducing downstream target genes.

In contrast to higher metazoans, where patterning mechanisms generally occur only during embryogenesis, *Hydra* tissue is in a state of constant growth and tissue replacement. Thus, axial patterning continuously maintains the different body parts⁶. Although *HyDsh* and *HyGSK3* are expressed uniformly at low levels throughout the animal in this dynamic tissue background (not shown), *in situ* hybridization revealed interesting expression patterns of *HyTcf* and *HyWnt*: expression was always restricted to the apical tip of the body axis in adult polyps (Fig. 2a, b). The *HyTcf* expression domain was about 200–300 μm wide, comprising the entire hypostome and exhibiting a gradient-like distribution (Fig. 2a). The *HyWnt* expression domain measured only about 60–100 μm in diameter, corresponding to about 50 epithelial cells (Fig. 2b). This apical tissue represents the *Hydra* head organizer^{7,8}.

New polyps are continually generated by an asexual budding process in the lower body column. During this process, significant upregulation of *Hy β -Cat* expression occurs in a ring-like domain in the prospective budding zone, just before tissue evagination starts (Fig. 2c). This high expression is maintained until bud detachment (Fig. 2d, e). *HyTcf* expression closely resembles *Hy β -Cat* expression, as it is induced in the same ring-like domain (Fig. 2f). Thereafter, expression of *HyTcf* is restricted to the side of the new bud protrusion (Fig. 2g) and finally to the hypostome (Fig. 2h). Temporal and spatial

co-expression of *Hy β -Cat* and *HyTcf* during bud initiation indicates that both may be involved in transcriptional regulation⁹. In addition, *Hy β -Cat* is expressed at a low level throughout the polyp (not shown), which might account for a second function in cell–cell adhesion. This view is supported by uniform expression of a *Hydra* Cadherin, which we identified by its interaction with *Hy β -Cat* in two-hybrid assays (not shown).

By comparison, *HyWnt* expression began in a few cells in the presumptive bud tissue just at the onset of tissue evagination (Fig. 2i; stage 1 in ref. 10). This spot increased in size, until it resembled the hypostomal expression domain of full-grown polyps (Fig. 2j, k). Together, these results indicate that a localized Wnt signalling centre may be established in a morphogenetic field (the budding zone) defined by elevated expression of *Hy β -Cat* and *HyTcf*.

Formation of a head organizer can also be induced by head regeneration. Removal of the head leads to morphallactic regeneration of a new head in the distal tissue of the remaining body column^{6,8}. Wound healing occurs within 30–60 min, and the head organizer is re-established between 2 and 8 h after head removal. Tentacle and hypostomal structures appear after 30–36 h. *Hy β -Cat* and *HyTcf* are upregulated in the regenerating tips within 1 h

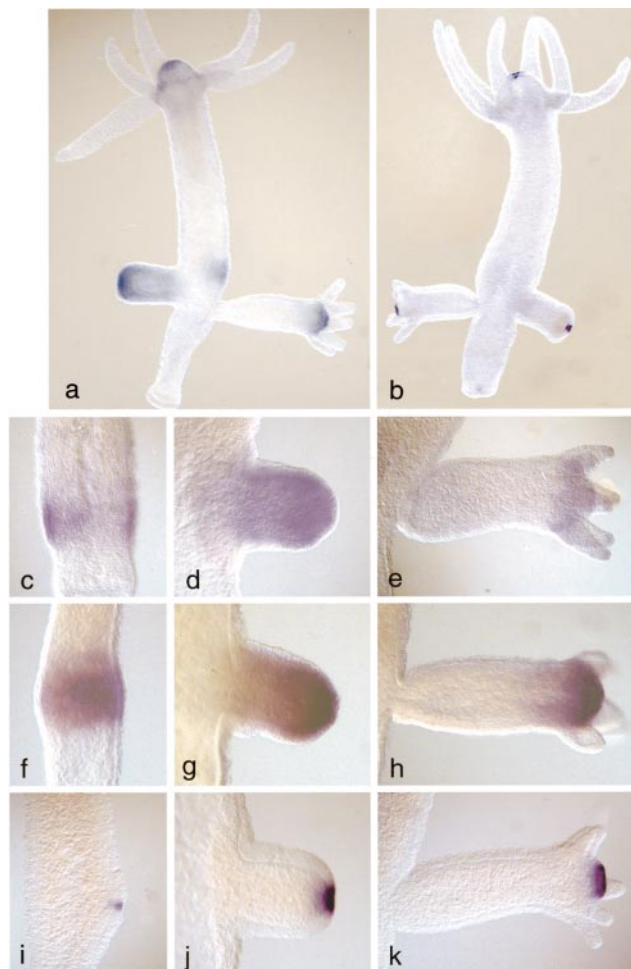


Figure 2 Expression of Wnt signalling molecules in morphogenesis of intact *Hydra* polyps visualized by whole-mount *in situ* hybridization. **a**, *HyTcf* is expressed in the head and the budding zone. **b**, *HyWnt* is expressed in the tips of the hypostome and evaginating buds. **c–k**, Expression dynamics during bud formation. **c–e**, *Hy β -Cat*; **f–h**, *HyTcf*; **i–k**, *HyWnt*. Bud stages according to ref. 10: **c**, **f**, nonbudding polyps; **i**, stage 1; **d**, **g**, **j**, stage 2; **e**, **h**, **k**, stage 9.

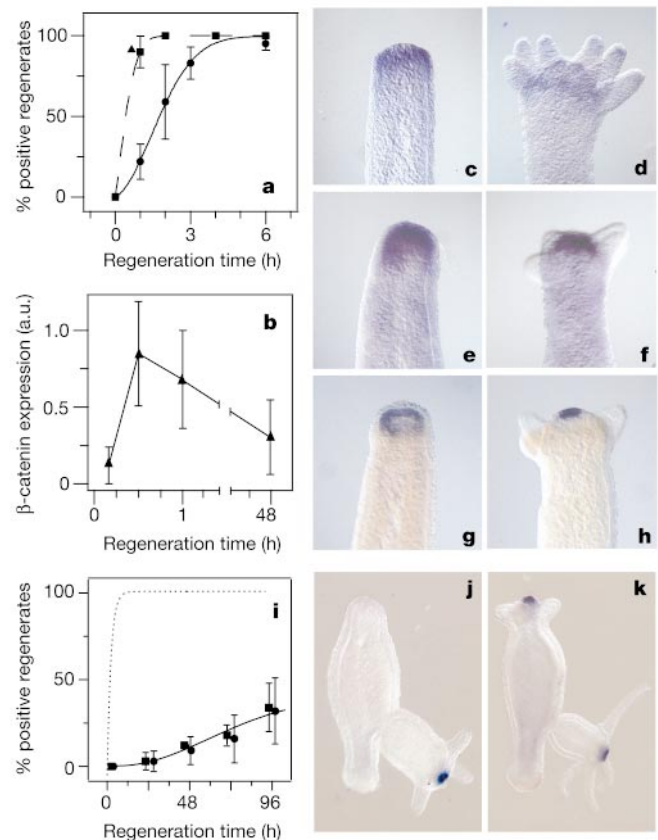


Figure 3 Expression of Wnt signalling molecules during head regeneration. **a**, Local upregulation of *Hy β -Cat*, *HyTcf* and *HyWnt* expression in head regenerating tips of wild-type *H. vulgaris*. *HyTcf* (squares) and *HyWnt* (circles) data represent mean ($\pm$ s.d.) of three independent experiments; $n = 10$ per data point in each experiment. The 1-h *Hy β -Cat* (triangle) value represents analysis of 37 regenerates. **b**, Upregulation of *Hy β -Cat* expression in the same experiment as assayed by RT–PCR. Data represent *Hy β -Cat*/*HyEF1 α* -ratios plotted as the mean ($\pm$ s.d.) of three independent experiments. **c–h**, Whole-mount *in situ* hybridizations at 1 h (**c**, **e**, **g**) and 48 h (**d**, **f**, **h**) after head removal. **c**, **d**, *Hy β -Cat*; **e**, **f**, *HyTcf*; **g**, **h**, *HyWnt*. **i**, Local upregulation of *HyTcf* and *HyWnt* expression in head regenerating tips of mutant *H. magnipapillata* strain reg-16. *HyTcf* (squares) and *HyWnt* (circles) data represent mean ($\pm$ s.d.) of three independent experiments; $n = 6$ –18 per data point in each experiment. For comparison, the *HyWnt* kinetics of wild-type regenerates (**a**) is indicated as dotted line. **j**, **k**, *HyWnt* expression in non-regenerating (**j**) and delayed regenerating (**k**) reg-16 polyps 96 h after head removal.

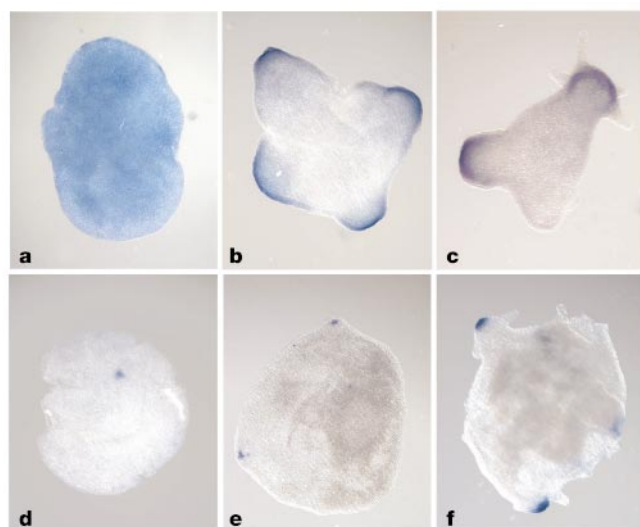


Figure 4 Expression dynamics of *HyTcf* and *HyWnt* during aggregate development. **a–c**, *HyTcf*; **d–f**, *HyWnt*. Aggregates were assayed by *in situ* hybridization at 24 h (**a**, **d**), 48 h (**b**, **e**) and 96 h (**c**, **f**).

of head removal (Fig. 3a, c, e). Quantitative PCR with reverse transcription (RT–PCR) experiments confirmed the extremely early upregulation of *Hyβ-Cat* expression (Fig. 3b). *HyWnt* expression could be observed in regenerating tips as early as 1 h after head removal and, after 6 h, all regenerates exhibited this expression pattern (Fig. 3a, g). The local transcriptional upregulation of *Hyβ-Cat* and *HyTcf* is an early event in regeneration of the head organizer, followed by local transcriptional activation of *HyWnt*. During the following 48 h, the expression patterns seen in steady-state heads are gradually re-established (Fig. 3d, f, h).

To obtain further evidence that Wnt signalling molecules are activated during head organizer formation, we analysed *HyWnt* and *HyTcf* expression in the regeneration-deficient mutant strain *H. magnipapillata* reg-16. Bud formation is unaffected in reg-16 mutants, but head regeneration is severely impaired. Most reg-16 polyps cannot regenerate a head, and in the remaining few regenerates head development is delayed¹¹. In contrast to wild-type regenerates (Fig. 3a), none of the reg-16 regenerates showed local upregulation of *HyTcf* and *HyWnt* expression earlier than 24 h after head removal (Fig. 3i). After 96 h, about 30% expressed *HyTcf* and *HyWnt* in their regenerating tips and these regenerates also started to form tentacles and a hypostome (Fig. 3k). Seventy per cent of the regenerates expressed neither *HyTcf* nor *HyWnt* and showed no morphological head structures (Fig. 3j). Thus, expression of *HyTcf* and *HyWnt* is tightly correlated with head formation in wild-type and mutant head regenerates.

A possible caveat for the instructive role of the Wnt pathway during bud formation and head regeneration comes from the fact that the polyp's body axis and its polarity remain intact. We therefore analysed aggregates from dissociated cell suspensions, where positional information is completely lost¹². Head organizers form *de novo* at random locations in the aggregates and induce surrounding tissues to develop new body axes, which finally separate into intact polyps¹². *HyTcf* was expressed in epithelial cells of the entire aggregate after 24 h (Fig. 4a). Uniform *HyTcf* expression then progressively separated into several distinct domains at the sites of new head organizing centres (Fig. 4b, c) indicating that *HyTcf* transcription is downregulated in all cells that fail to differentiate into head-specific tissue. In the same experiment, small spots of about 10–20 *HyWnt*-expressing cells formed as early as 24 h after the start of aggregation (Fig. 4d). At this time, epithelial cell sorting into ectodermal and endodermal cell layers is complete¹², indicating that formation of intact tissues may be a

prerequisite for *HyWnt* expression. These *HyWnt*-positive spots clearly localized with head protrusions at later stages (Fig. 4e, f), and quantitative analysis revealed that each *HyWnt*-positive spot finally developed into a head (not shown). Thus, *HyWnt* and *HyTcf* seem to be early components of the molecular network, setting up *de novo* a small number of cells with a strong axis-inducing capacity.

Reaction-diffusion mechanisms of local self-activation and long-range inhibition have been postulated to generate gradients of diffusible morphogens *de novo*^{8,13}. We propose that *HyWnt* signalling is involved in local self-activation in the *Hydra* head organizer. Two lines of evidence support this view. First, the timing and localization of *HyWnt* expression during head regeneration correlate strongly with the local increase in head activation as measured in classical transplantation experiments¹⁴. Second, in *Drosophila*, induction of armadillo/Tcf transcription complexes by wingless (*wg*)¹⁵ and identification of a Tcf-binding site in the *wg* promoter¹⁶ suggest direct *wg* self-activation. Our findings that during head regeneration and budding *Hyβ-Cat* and *HyTcf* are transcriptionally upregulated significantly earlier than *HyWnt* are consistent with the idea that *HyWnt* is also a downstream target for *Hyβ-Cat*/*HyTcf* complexes.

Our results also have implications for the evolution of the metazoan body plan. They indicate that a Wnt signalling centre involved in axis formation may have existed in the common ancestor of Cnidaria and bilateral metazoans, and thus originated in the earliest multicellular animals. Cnidaria with one major body axis have been identified from late proterozoic, ediacaran sediments and are ancestral to the diversification of bilateral protostomes and deuterostomes^{17–19}. An earlier model of the molecular nature of bilaterian ancestors argues that HOX genes constitute the metazoan anterior–posterior (A–P) body axis²⁰. In fact, orthologues of a number of HOX genes have been identified in the Cnidaria, and the existence of ProtoHox clusters has been proposed^{21–23}. Sequential patterns of axial expression are, however, yet not clearly demonstrated.

As Wnt signalling cascades act in the establishment of dorsal–ventral (D–V) polarity and D–V axis formation in *Drosophila* and vertebrates^{24–26}, our findings that Wnt signalling is involved in axis formation in *Hydra* open the following perspectives. (1) Both signalling cascades of the D–V patterning system and a primitive HOX cluster are present in the Cnidaria and fulfill a synergistic function in axis formation. According to that view, A–P and D–V patterning mechanisms existed in early, radially symmetric metazoans and became spatially uncoupled during bilaterian evolution. (2) Alternatively, the cnidarian body axis is related to the D–V axis of higher metazoans, implying that the A–P axis evolved at the transition from radial to bilateral animals, as already postulated by classical morphological comparisons^{27,28}. Further study of cnidarians will give insight into the evolution of the A–P and D–V axes of bilaterians. □

Methods

Cloning

Nested PCR amplifications were done with a single-stranded complementary DNA template primed from 250 ng of *Hydra vulgaris* polyA⁺ RNA. The primer combinations were as follows: Wnt1/Wnt5→Wnt1/Wnt3, Dsh1/Dsh4→Dsh1/Dsh3, GSK6/GSK5→GSK6/GSK3. Degenerated primer sequences (IUB code; 5'→3'): Wnt1: TGGSANTGGGGNGGNTG; Wnt3: CCNGCNYBRTTRTTRTG; Wnt5: TYNCCRTGRCAYTTTCA; Dsh1: GGNMGNATHGARCNGG; Dsh3: GGRTCCCARCAYTTNGC; Dsh4: GTDATYTTNARCCACAT; GSK3: GGYTTDATTCNCCKRTGCA; GSK5: TTNGCWSWNCRCARATCRCA; GSK6: GTNGCNATHAARAARGT. PCR conditions: 3 min 95°C (1 cycle); 1 min 95°C, 1 min 42°C, 1 min 72°C (30 cycles); 2 min 72°C (1 cycle). We cloned and sequenced amplification products of the expected size using standard procedures. To obtain the translational open reading frame (ORF) of *HyWnt* and *HyDsh*, 5' and 3' rapid amplification of cloned ends (RACE) experiments were carried out. 5' RACE was performed using a cDNA produced with the SMART system (Clontech), whereas 3' RACE was done according to ref. 29. Using the *Hydra*-specific GSK6/GSK3 PCR fragment as a probe, a full-length *HyGSK* cDNA clone was obtained by screening a *Hydra vulgaris* cDNA library (Stratagene).

Yeast two-hybrid screening

We performed yeast two-hybrid experiments with a *Hydra vulgaris* two-hybrid cDNA library (Stratagene) as described by the supplier. The *Hyβ-Cat* bait (amino acids 166–586) was cloned to the GAL4 DNA-binding domain in the pBD phagemid vector. The bait was transformed into the yeast strain YRG-2 together with the cDNA library fused to the GAL4 activation domain in the pAD phagemid vector. Interacting proteins were selected on plates lacking histidine and containing 3-aminotriazole (10 mM). We tested activation of a second reporter gene, lacZ, by filter lift assay.

Xenopus embryo injection

Xenopus embryos were generated by *in vitro* fertilization, dejellied and further cultivated. Dorsal–ventral polarity was determined at the four-cell stage, and darker, ventral blastomeres were injected with 4 nl synthetic mRNA (2–4 ng). For mRNA synthesis, linearized pBS-plasmid containing full-length *Hyβ-Cat* cDNA was transcribed using a T3 Message Machine Kit (Ambion).

In situ hybridization

Whole mount *in situ* hybridization with DIG-labelled RNA probes was done as described³⁰.

RT-PCR

A cDNA template was produced using Ready-To-Go First-Strand Beads (Pharmacia) from total RNA isolated from the distal fifth of head regenerates. We detected *Hyβ-Cat* using the following primers: *β-Cat* forward (5'-CATTCTCGAAATGATGGC-3') and *β-Cat* reverse (5'-CGCGTGGTAAAAACACA-3'). A 75-base pair (bp) intron located between these two primer sites in the *Hyβ-Cat* gene served as a control for genomic contamination. *Hydra EF1α* (GenBank accession number: Z68181) was detected using the primers EF1 forward (5'-GTTTGAAGCTGGTATTTC-3') and EF1 reverse (5'-TGTTTACCAGTATCTTG-3'). PCR conditions: 5 min 95 °C (1 cycle); 1 min 95 °C, 1 min 46 °C, 1 min 72 °C (35 cycles); 5 min 72 °C (1 cycle). Amplification products were quantified by optical density analysis (Image Master VDS, Pharmacia).

Hydra tissue manipulations

Hydra mass culture was maintained in a daily feeding regime³⁰. All experimental animals were starved for 24 h. For head regeneration, polyps were bisected at 75% body length (one-quarter below the hypostome) and the proximal pieces were immediately transferred into fresh culture medium for further head regeneration. For dissociation and reaggregation, heads and feet were removed from 300 animals. The remaining body columns were then mechanically dissociated in hypotonic dissociation medium and reaggregated¹².

Received 24 February; accepted 20 July 2000.

- Cadigan, K. M. & Nusse, R. Wnt signalling: a common theme in animal development. *Genes Dev.* **11**, 3286–3305 (1997).
- Cox, R. T. & Peifer, M. Wingless signalling: The inconvenient complexities of life. *Curr. Biol.* **8**, R140–R144 (1998).
- Hobmayer, E. *et al.* Identification of a hydra homologue of the β -catenin/plakoglobin/armadillo gene family. *Gene* **172**, 155–159 (1996).
- Minobe, S. *et al.* Identification and characterization of the epithelial polarity receptor "Frizzled" in *Hydra vulgaris*. *Dev. Genes Evol.* **210**, 258–262 (2000).
- Molenaar, M. *et al.* XTCF-3 transcription factor mediates β -catenin-induced axis formation in *Xenopus* embryos. *Cell* **86**, 391–399 (1996).
- Bode, P. M. & Bode, H. R. in *Pattern Formation. A Primer in Developmental Biology* (eds Malacinski, G. M. & Bryant, S. V.) 213–241 (Macmillan, New York, 1984).
- Meinhardt, H. A model for pattern formation of hypostome, tentacles, and foot in hydra: How to form structures close to each other, how to form them at a distance. *Dev. Biol.* **157**, 321–333 (1993).
- Wolpert, L. *Principles of Development* (Oxford Univ. Press, Oxford, 1998).
- Niehrs, C. & Pollet, N. Synexpression groups in eukaryotes. *Nature* **402**, 483–487 (1999).
- Otto, J. J. & Campbell, R. D. Budding in *Hydra attenuata*: Bud stages and fate map. *J. Exp. Zool.* **200**, 417–428 (1977).
- Achermann, J. & Sugiyama, T. Genetic analysis of developmental mechanisms in Hydra. X. Morphogenetic potentials of a regeneration-deficient strain (reg-16). *Dev. Biol.* **107**, 13–27 (1985).
- Gierer, A. *et al.* Regeneration of hydra from reaggregated cells. *Nature New Biol.* **239**, 98–101 (1972).
- Gierer, A. & Meinhardt, H. A theory of biological pattern formation. *Kybernetik* **12**, 30–39 (1972).
- MacWilliams, H. K. *Hydra* transplantation phenomena and the mechanism of *Hydra* head regeneration. II. Properties of head activation. *Dev. Biol.* **96**, 239–257 (1983).
- van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene *DTCF*. *Cell* **88**, 789–799 (1997).
- Lessing, D. & Nusse, R. Expression of wingless in the *Drosophila* embryo: a conserved *cis*-acting element lacking conserved *Ci*-binding sites is required for patched-mediated repression. *Development* **125**, 1469–1476 (1998).
- Conway Morris, S. Metazoan phylogenies: falling into place or falling to pieces? A palaeontological perspective. *Curr. Opin. Genet. Dev.* **8**, 662–667 (1998).
- Aguinaldo, A. M. A. *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**, 489–493 (1997).
- Knoll, A. H. & Carroll, S. B. Early animal evolution: Emerging views from comparative biology and geology. *Science* **284**, 2129–2137 (1999).
- Slack, J. M. W., Holland, P. W. H. & Graham, C. F. The zootype and the phylotypic stage. *Nature* **361**, 490–492 (1993).
- Finnerty, J. R. & Martindale, M. Q. The evolution of the Hox cluster: insights from outgroups. *Curr. Opin. Genet. Dev.* **8**, 681–687 (1998).

- Kuhn, K., Streit, B. & Schierwater, B. Isolation of Hox genes from the scyphozoan *Cassiopeia xamachana*: Implications for the early evolution of Hox genes. *Mol. Dev. Evol.* **285**, 63–75 (1999).
- Gauchat, D. *et al.* Evolution of Antp-class genes and differential expression of *Hydra Hox/paraHox* genes in anterior patterning. *Proc. Natl Acad. Sci. USA* **97**, 4493–4498 (2000).
- Neumann, C. J. & Cohen, S. M. Long-range action of Wingless organizes the dorsal–ventral axis of the *Drosophila* wing. *Development* **124**, 871–880 (1997).
- Kopp, A., Blackman, R. K. & Duncan, I. Wingless, Decapentaplegic and EGF receptor signalling pathways interact to specify dorso–ventral pattern in the adult abdomen of *Drosophila*. *Development* **126**, 3495–3507 (1999).
- Larabell, C. A. *et al.* Establishment of the dorso–ventral axis in *Xenopus* embryos is presaged by early asymmetries in β -catenin that are modulated by the Wnt signalling pathway. *J. Cell Biol.* **136**, 1123–1136 (1997).
- Willmer, P. *Invertebrate Relationships. Patterns in Animal Evolution* (Cambridge Univ. Press, Cambridge, UK, 1990).
- Nielsen, C. *Animal Evolution: Interrelationships of the Living Phyla* (Oxford Univ. Press, Oxford, 1995).
- Frohman, M. A. in *PCR Primer—A Laboratory Manual* (eds Dieffenbach, C. W. & Dveksler, G. S.) 381–409 (CSHL, New York, 1995).
- Technau, U. & Bode, H. R. *HyBra1*, a brachyury homologue, acts during head formation in hydra. *Development* **126**, 999–1010 (1999).

Supplementary information is available on Nature's Word-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Sarraz for sharing unpublished data, the Mishima *Hydra* lab for providing reg-16 mutants, C. N. David and U. Technau for critical comments on the manuscript, and K. Wehner and P. L  bberich for assistance in cloning. This work was supported by grants from the Deutsche Forschungsgemeinschaft (B.H. and T.W.H.).

Correspondence and requests for materials should be addressed to B.H. (e-mail: hobmayer@bio.tu-darmstadt.de) or T.W.H. (e-mail: holstein@bio.tu-darmstadt.de). GenBank accession numbers: *HyWnt* (AF272673), *HyDsh* (AF272674), *HyGSK3* (AF272672) and *HyTcf* (AF271696).

Identification of a vesicular glutamate transporter that defines a glutamatergic phenotype in neurons

Shigeo Takamori*, Jeong Seop Rhee†, Christian Rosenmund† & Reinhard Jahn*

* Department of Neurobiology and † Department of Membrane Biophysics, Max-Planck-Institute for Biophysical Chemistry, Am Fa  berg 11, D-37077 G  ttingen, Germany

Glutamate is the major excitatory neurotransmitter in the mammalian central nervous system. Synaptic vesicles are loaded with neurotransmitter by means of specific vesicular transporters. Here we show that expression of BNPI, a vesicle-bound transporter associated with sodium-dependent phosphate transport^{1–3}, results in glutamate uptake by intracellular vesicles. Substrate specificity and energy dependence are very similar to glutamate uptake by synaptic vesicles. Stimulation of exocytosis—fusion of the vesicles with the cell membrane and release of their contents—resulted in quantal release of glutamate from BNPI-expressing cells. Furthermore, we expressed BNPI in neurons containing GABA (γ -aminobutyric acid) and maintained them as cultures of single, isolated neurons that form synapses to themselves. After stimulation of these neurons, a component of the postsynaptic current is mediated by glutamate as it is blocked by a combination of the glutamate receptor antagonists, but is insensitive to a GABA_A receptor antagonist. We conclude that BNPI functions as vesicular glutamate transporter and that expression of BNPI suffices to define a glutamatergic phenotype in neurons.

Specific vesicular transporters are responsible for loading synaptic vesicles with neurotransmitters. Unlike transporters at the plasma membrane, these are driven by an electrochemical proton gradient across the vesicle membrane. Proteins responsible for

The vesicular glutamate transporter is known to prefer glutamate over aspartate, in contrast to the Na^+ -dependent transport systems⁹.

a

	Homogenate	P1	S2	P2	LP1	LP2	Peak I	CPG-column SV
BNPI	~55	~55	~55	~55	~55	~55	~55	~55
VGAT	~55	~55	~55	~55	~55	~55	~55	~55
Synaptophysin	~55	~55	~55	~55	~55	~55	~55	~55
NMDA receptor	~55	~55	~55	~55	~55	~55	~55	~55

b

	BNPI Beads		VGAT Beads		Control Beads	
	Input	Sup	Beads	Sup	Beads	Sup
BNPI	~55	~55	~55	~55	~55	~55
VGAT	~55	~55	~55	~55	~55	~55
VMAT2	~55	~55	~55	~55	~55	~55
pp116 subunit	~55	~55	~55	~55	~55	~55
Synaptophysin	~55	~55	~55	~55	~55	~55

c

Glutamate uptake (c.p.m.)

Condition	Glutamate uptake (c.p.m.)
BNPI	~3,600
VGAT	~800
Cont	~1,200

GABA uptake (c.p.m.)

Condition	GABA uptake (c.p.m.)
BNPI	~250
VGAT	~520
Cont	~120

190

the membrane-potential component of the proton electrochemical potential across the vesicle membrane, which is generated by a vacuolar proton ATPase¹⁰. Bafilomycin A₁ and *N*-ethylmaleimide (NEM) both specific inhibitors of vacuolar proton pumps, inhibited glutamate uptake, whereas inhibitors of mitochondrial F₀-F₁-ATPases and E₁-E₂-ATPases were ineffective (Fig. 2e). In addition, conditions known to reduce the membrane-potential component of the proton gradient (high chloride, electrogenic ionophore combinations) inhibited uptake, whereas a selective abolishment of the pH gradient (ammonium sulphate, nigericin) had only minor effects (Fig. 2f), in agreement with previous results on synaptic vesicles^{11,12}.

The results described above show that BNPI expression in BON cells leads to glutamate uptake in a vesicular compartment that is acidified by a vacuolar ATPase. BON cells contain secretory vesicles that can undergo regulated exocytosis in response to muscarinic stimulation⁸. We therefore asked whether BNPI expression would lead to accumulation of glutamate in secretory vesicles in addition to serotonin, and, consequently, whether glutamate is released upon stimulation in discrete quanta. To analyse for glutamate release, an excess of BNPI-expressing BON cells was co-cultured with HEK293 cells that were transiently transfected with a non-desensitizing

variant of the AMPA receptor¹³. To identify the BNPI- and glutamate-receptor (GluR)-expressing cells, green and red fluorescent proteins (EGFP and Ds-Red, respectively) were co-expressed using bi-cistronic constructs.

Whole-cell patch-clamp measurements were performed from GluR cells and their responsiveness to glutamate was tested by application of saturating concentration of L-glutamate (1 mM). Cells with robust responses (> 3 nA) were then placed in direct contact to BNPI cells. When exocytosis of the BNPI cells was stimulated by acetylcholine, transient miniature excitatory post-synaptic current (mEPSC)-like inward currents frequently developed in the GluR cells (14 out of 18 experiments; Fig. 3). These events were observed with an average frequency of 0.15 Hz. They produced inward currents with a peak amplitude of 51 ± 4 pA, a rise time of 4.1 ± 1.1 ms and a decay time constant of 24 ± 5 ms ($n = 275$, Fig. 3d). When the specific AMPA-receptor antagonist NBQX (30 μ M, Fig. 3a) was added, the response was largely inhibited: spike-like events occurred with a 23-fold lower frequency (10 events in 1,200 s), and had an average amplitude of 22 ± 6 pA (false-positive events). No significant NBQX-sensitive activity could be recorded from VGAT-expressing cells regardless of whether NBQX was present or not ($n = 7$; Fig. 3d).

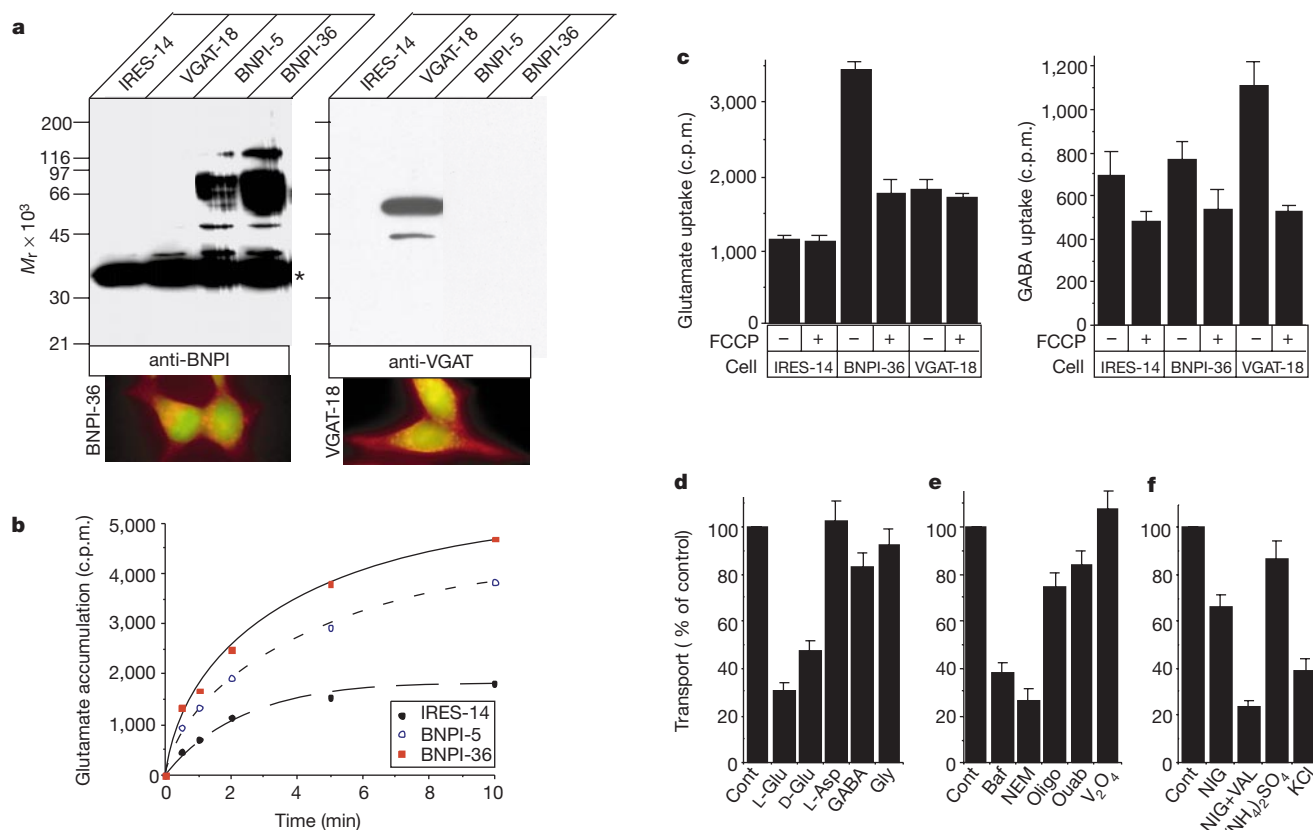


Figure 2 BNPI functions as a vesicular glutamate transporter. **a**, Intracellular expression of BNPI and VGAT in BON stable transfectants monitored by immunoblotting (top) and immunofluorescence (bottom, red). Asterisk, an unrelated immunoreactive band that is present in all cell clones (see also Fig. 1). Note that all stable transfectants express EGFP (bottom, green). **b**, Time dependence of [³H]-glutamate accumulation of membranes isolated from BNPI-expressing clones (BNPI-5 and BNPI-36) compared with that from control clones (IRES-14). **c**, [³H]-glutamate uptake by membranes from BNPI-expressing cells (BNPI-36) is sensitive to the uncoupler FCCP (left), whereas no such uptake is present in membranes from the other cell lines. By contrast, enhanced FCCP-sensitive [³H]GABA uptake is observed in VGAT-expressing clone (VGAT-18). **d**, Substrate specificity of glutamate uptake in BNPI-expressing cell membranes, measured by

competition of [³H]-glutamate uptake with 10 mM unlabelled substrate. **e**, Glutamate uptake in BNPI-expressing cell membranes is driven by the vacuolar ATPase as it is inhibited by bafilomycin A₁ (Baf; 100 nM) and *N*-ethylmaleimide (NEM; 0.2 mM), but not by oligomycin B (Oligo; 5 μ M), ouabain (Oub; 2 mM) or vanadate (V₂O₄; 50 μ M). **f**, Glutamate uptake in BNPI-expressing cells depends more on the membrane potential than on the pH gradient, as shown by selective sensitivity towards specific inhibitors. The following concentrations were used: nigericin (NIG; 5 μ M), valinomycin (VAL; 20 μ M), (NH₄)₂SO₄ (10 mM), and KCl (100 mM). In **b**, **c**, background uptake was determined by incubation at 4 °C for 10 min. In **d**–**f**, values are expressed as percentage of control (without additives). Error bars, s.e.m. from at least two independent experiments using different membrane preparations.

Finally, we investigated whether expression of BNPI in neurons suffices to cause synaptic release of glutamate. For this purpose, autaptic hippocampal neurons were infected with a semliki-forest virus construct co-expressing BNPI and EGFP. EGFP-positive GABA-releasing neurons were selected for whole-cell recordings. In all cases we ensured that these cells were strictly isolated from other neurons. Autaptic responses were evoked by brief somatic depolarization (-70 mV to 0 mV for 1 – 2 ms, Fig. 4). All cells exhibited a robust inhibitory postsynaptic current (IPSC, 4.3 ± 1.7 nA, $n = 9$). In uninfected cells, this current (4.2 ± 0.7 nA) was completely inhibited by the GABA_A-receptor antagonist bicuculline and picrotoxin ($n = 13$). In BNPI-expressing cells, however, we regularly (9 of 12 cells) observed a fast inward current component that was resistant to saturating concentrations of the bicuculline. This current component had an average amplitude of 208 ± 32 pA ($n = 9$) and was identified as a glutamate-mediated excitatory postsynaptic current (EPSC) as it was blocked by a combination of the AMPA-receptor and NMDA-receptor antagonists NBQX and AP5, respectively (Fig. 4d). The EPSC kinetics were slightly slower in rise and decay than expected from a normal glutamate-mediated EPSC (Fig. 4e, f). This is probably due to the

fact that glutamate concentrations are lower in vesicles that contain both glutamate and GABA than in pure glutamate-containing vesicles, leading to lower receptor occupancy and activation kinetics. Alternatively, the activated AMPA receptors reside outside of the active zone of the otherwise GABA-mediated autapses. This result shows that exogenous expression of BNPI in GABA-containing neurons results in a hybrid phenotype, effectively releasing both the excitatory transmitter glutamate and the inhibitory transmitter GABA from the same neuron.

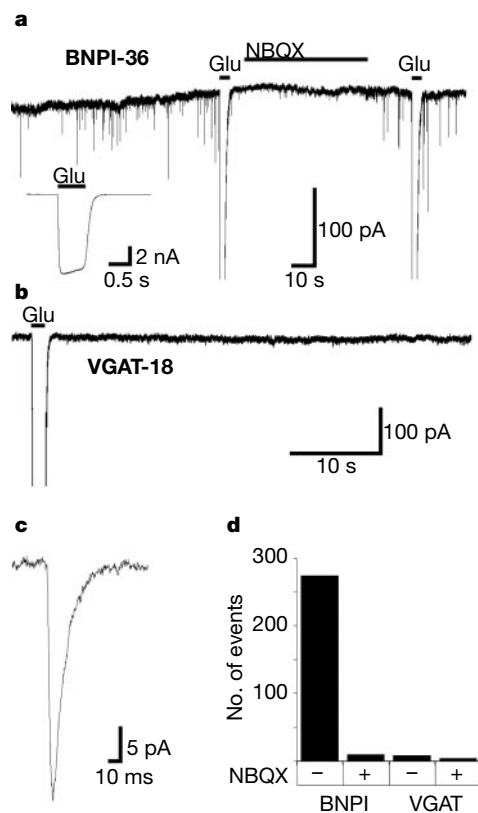


Figure 3 Quantal release of glutamate from BNPI-expressing BDN cells detected by AMPA-receptor-expressing reporter cells. **a**, HEK293 cells expressing a non-desensitizing AMPA receptor were whole-cell voltage clamped at -70 mV and manipulated into proximity with BDN cells expressing BNPI. Acetylcholine was applied at 10 – 100 μ M to induce secretion from BDN cells. The sensitivity of the HEK293 cell to glutamate was intermittently tested by application of exogenous L-glutamate (1 mM, black bars; also see inset). Spike-like inward currents were observed that were stimulated 2.8 -fold on acetylcholine application. These events were blocked by NBQX and reversed direction on depolarization to positive potentials (not shown). **b**, An example recording from a HEK/VGAT-18 pair under identical experimental conditions as shown in **a**. Acetylcholine, but not NBQX was added to the external medium. **c**, Average response from 16 events from the same cell as shown in **a**. **d**, Event statistics in absence and presence of NBQX. A total of $1,200$ s recording time from 14 experiments on BNPI-36 cells and 900 s from 7 experiments in VGAT-18 cells were analysed.

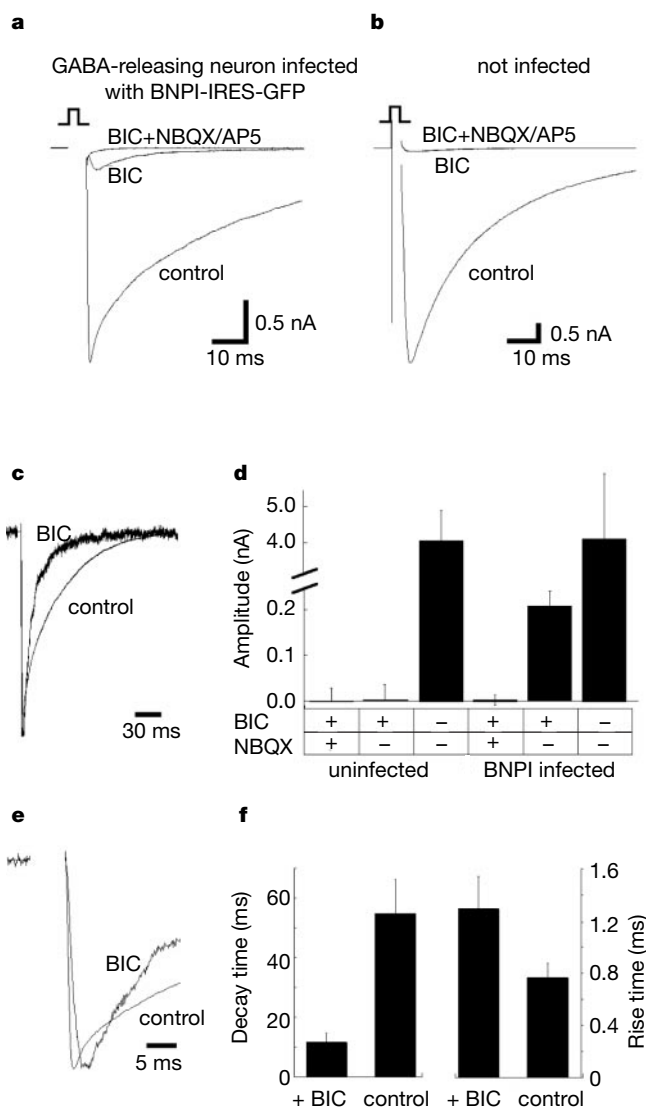


Figure 4 Co-release of GABA and glutamate from BNPI-expressing GABA-containing hippocampal neurons in autaptic culture. **a**, Autaptic response from a hippocampal GABA-containing neuron infected with BNPI-SFV virus. Shown are three superimposed IPSCs, one in absence of blockers showed a typical robust slowly decaying inward current (control). This response was not completely blocked by saturating concentrations of bicuculline (BIC, 30 μ M) and picrotoxin (100 μ M). The remaining component was blocked by addition of NBQX (30 μ M) and AP-5 (50 μ M). **b**, Three superimposed IPSCs of an uninfected GABA-containing neuron. The current under control conditions was blocked by bicuculline and picrotoxin (BIC). Addition of NBQX and AP5 has no further effect on the block. **c**, The current transients in BIC+NBQX+AP5 were subtracted from the two other current components taken from **a**, and the inward currents were normalized to illustrate the time course of the two current components. **d**, Statistics of control, BIC and BIC+NBQX amplitudes from BNPI-infected and uninfected GABA-containing neurons. **e**, Expanded time axis of **c**. The evoked NBQX-sensitive current showed a delayed onset. **f**, Statistics of current-decay and current-rise time constants for evoked responses under control and bicuculline conditions. The intracellular solution contained 120 mM Cl^- .

Together, the data described above provide conclusive evidence that BNPI functions as vesicular glutamate transporter with properties identical to those observed for glutamate uptake by isolated synaptic vesicles. In fact, our data show that no other component is needed to make a neuron store glutamate in synaptic vesicles and release it by exocytosis upon stimulation. Thus, the vesicular glutamate transporter is the only specific component that is responsible for the transmitter selectivity of glutamate-releasing neurons and thus sets them apart from neurons containing other neurotransmitters. However, the involvement of Na^+ -dependent phosphate transport activity observed upon expression of BNPI (refs 1, 2) needs to be clarified. We cannot exclude that glutamate and phosphate transport are somehow linked, for example, that the transporter possesses a mode in which glutamate is exchanged for phosphate (in agreement with the opposite orientation of the two transport activities) with sodium providing some charge compensation.

Is BNPI the only vesicular glutamate transporter? In *C. elegans*, loss-of-function mutation in the BNPI orthologue eat-4 leads to impairment but not to a complete loss of glutamate-mediated neurotransmission^{14,15}. Such loss would be expected when the vesicles are not filled with glutamate. However, database searches yielded at least five *C. elegans* homologues of BNPI (accession numbers S40767, Q10046, T23589, Q03567 and S28286), one of which is closely related to eat-4 (S40767), indicating that these additional molecules may operate alongside eat-4 as a vesicular glutamate transporter. Similarly, a human complementary DNA encoding a close relative of BNPI (DNPI, 80% sequence identity with BNPI) has recently been isolated¹⁶. Thus, it appears that BNPI represents a small family of conserved vesicular glutamate transporters that are, at least in part, functionally redundant. □

Methods

Plasmids and antibodies

The cDNAs corresponding to the open reading frame of rat BNPI and rat VGAT were amplified by polymerase chain reaction using as a template cDNA that was reverse-transcribed from rat brain total RNA, subcloned into pcDNA3.1 (Invitrogen) at *HindIII*–*XhoI* and *BamHI*–*XbaI* sites, respectively, and the sequences were confirmed. The 1.7 kilobases (kb) of *NheI*–*XhoI* fragment from BNPI–pcDNA3.1 and the 1.6 kb of *PmeI* fragment from VGAT–pcDNA3.1 were further subcloned into pIRES2–EGFP (Clontech) at *NheI*–*XhoI* and *SmaI* site, respectively. For the Semliki-forest virus (SFV) construct for BNPI, 2.9 kb of *PmeI*–*DraI* fragment from BNPI–pIRES2–EGFP, which contains open reading frame of BNPI and the internal ribosome entry site (IRES) followed by EGFP were cloned into pSFV1 (Gibco BRL) at *SmaI* site. Direction of the insert was confirmed by restriction mapping. The virus was prepared as described¹⁷. A rabbit polyclonal antibody directed against the amino-terminal part of rat BNPI was generated by injecting the synthetic peptide (EKRQEGAETLELSAD-C; residues 24–38) which corresponds to the predicted amino-acid sequence of rat BNPI (ref. 1). Other antibodies used in this study are described elsewhere⁷.

BON stable transfectants

BON cell line, human pancreatic tumour cells, were cultured in Dulbecco's MEM/Nut mix F-12 (1:1), supplemented with 10% fetal bovine serum, 100 IU ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin. BNPI–pIRES2–EGFP, VGAT–pIRES2–EGFP or pIRES2–EGFP without insert was used for transfecting cells (20 µg) by means of the calcium-phosphate method¹⁸. The transfectant cells were then selected in the presence of 800 µg ml⁻¹ of G418 and the resulting clones were screened for EGFP fluorescence by fluorescence microscopy. The protein expression of BNPI and VGAT in resulting clones were further confirmed by both immunofluorescence and immunoblotting by using specific antibodies.

Membrane preparation from BON cells and neurotransmitter uptake assay

BON stable clones cultured on 15-cm dishes were washed twice with ice-cold PBS and then collected in 0.32 M sucrose, 4 mM HEPES–NaOH, pH 7.4. The cells were then homogenized using a ball-bearing homogenizer (10 µm clearance). The resulting homogenate was cleared by centrifugation at 10,000g for 5 min to remove the nuclei and cell debris. The supernatant was sedimented by centrifugation at 200,000g for 20 min in TLA120.2 rotor. The resulting membrane pellet was resuspended in 0.32 M sucrose, 4 mM KCl, 4 mM MgSO₄, 10 mM HEPES–KOH, pH 7.4 (uptake assay buffer). Neurotransmitter uptake assays were carried out as described previously with slight modifications^{7,19}. The reactions were started by incubating 50 µg membrane fraction with 40 µM glutamate or GABA containing 2 µCi [³H]glutamate or [³H]GABA (both from NEM) in the presence of 2 mM

ATP. Proton uncoupler, FCCP (final concentration, 46 µM) and a variety of additives were added when indicated.

Neuronal cell culture and electrophysiology

HEK293 cells were transfected with non-desensitizing AMPA-receptors (GluR1L497Y or the GluR2 homologue GluR2QL504Y; ref. 13) using standard procedures¹⁸. After 24–72 h after transfection, the GluR-expressing cells were identified either by co-transfecting pDsRed1-N1 (Clontech) or by employing a bicistronic pcDNA3 vector with an additional IRES–DsRed insert. BON cells expressing either BNPI or VGAT and EGFP as a marker were added at a density so that they reached 70% confluency in 24 h when experiments were performed. Spike-like inward currents were not counted if they had rise and decay time constants that were not consistent with receptor kinetics of the non-desensitizing GluR1/2. Hippocampal mouse neurons were cultured on microislands using standard procedures²⁰ and plated at low density (500 cells cm⁻²). The neuronal medium was serum-free (neurobasal medium A+B27, Gibco BRL). Experiments from inhibitory neurons were performed at 10–20 days *in vitro* and 24–27 h after addition of the SFV virus. The standard extracellular medium contained (in mM): NaCl, 140; KCl, 2.4; HEPES, 10; glucose, 10; CaCl₂, 4; MgCl₂, 4. The osmolarity was 300 mOsm and the pH was 7.3.

Solutions were applied using an array of quartz flow pipes positioned within 100–200 µm of the neuron, and connected to gravity-fed reservoirs. Each flow pipe was controlled by solenoid valves and was moved with a piezoelectric device under the control of computer software²¹. Patch pipette (borosilicate) resistance was 2–3.5 MΩ. Pipette solutions for neurons included (in mM): KCl, 120; HEPES, 10; EGTA, 1; MgCl₂, 4.6; Na₂ATP, 4; Creatinephosphate, 15; creatinephosphokinase 50 U ml⁻¹. Pipette solutions for HEK cells included (in mM) CsF, 140; HEPES, 10; EGTA, 10; NaCl, 10. The pH was 7.3 and the osmolarity was 300 mOsm. Currents were recorded using an Axopatch 200A or 200B amplifier (Axon). Series resistance was 60–90% compensated as was cell capacitance (5–25 pF). Electrophysiological data were acquired on an IBM 586 clone (Pclamp 7-8, Axon) and analysed on a Macintosh (AXOGRAPH4, Axon). The acquisition rate was 2–10 kHz, and data were filtered at 1–2 kHz. Data are expressed as means ± s.e.m.

Others

Purification of synaptic vesicles from rat brain, immuno-isolation and immunoblot analyses were performed as described^{7,22}. For peptide microsequencing, protein spots, neighbouring that of synaptotagmin, from several Commassie-stained gels were pooled, concentrated by Funnel-Well SDS–polyacrylamide gel electrophoresis²³ and digested with trypsin *in situ*. The resulting tryptic fragments eluted from the gel matrix were purified using reverse-phase high-performance liquid chromatography and subjected to laser desorption mass spectrometry before microsequencing⁵.

Received 18 July; accepted 21 August 2000.

- Ni, B., Rostek, P. R. Jr., Nadi, N. S. & Paul, S. M. Cloning and expression of a cDNA encoding a brain-specific Na⁺-dependent inorganic phosphate cotransporter. *Proc. Natl Acad. Sci. USA* **91**, 5607–5611 (1994).
- Ni, B. *et al.* Molecular cloning, expression, and chromosomal localization of a human brain-specific Na⁺-dependent inorganic phosphate cotransporter. *J. Neurochem.* **66**, 2227–2238 (1996).
- Bellochio, E. E. *et al.* The localization of the brain-specific inorganic phosphate transporter suggests a specific presynaptic role in glutamatergic transmission. *J. Neurosci.* **18**, 8648–8659 (1998).
- Reimer, R. J., Fon, E. A. & Edwards, R. H. Vesicular neurotransmitter transport and the presynaptic regulation of quantal size. *Curr. Opin. Neurobiol.* **8**, 405–412 (1998).
- Harteringer, J., Stenius, K., Hogemann, D. & Jahn, R. 16-BAC/SDS-PAGE: a two-dimensional gel electrophoresis system suitable for the separation of integral membrane proteins. *Anal. Biochem.* **240**, 126–133 (1996).
- Dent, J. A., Davis, M. W. & Avery, L. avr-15 encodes a chloride channel subunit that mediates inhibitory glutamatergic neurotransmission and ivermectin sensitivity in *Caenorhabditis elegans*. *EMBO J.* **16**, 5867–5879 (1997).
- Takamori, S., Riedel, D. & Jahn, R. Immunolocalization of GABA-specific synaptic vesicles defines a functionally distinct subset of synaptic vesicles. *J. Neurosci.* **20**, 4904–4911 (2000).
- Parekh, D. *et al.* Characterization of a human pancreatic carcinoid *in vitro*: morphology, amine and peptide storage, and secretion. *Pancreas* **9**, 83–90 (1994).
- Naito, S. & Ueda, T. Adenosine triphosphate-dependent uptake of glutamate into protein I-associated synaptic vesicles. *J. Biol. Chem.* **258**, 696–699 (1983).
- Maycox, P. R., Hell, J. W. & Jahn, R. Amino acid neurotransmission: spotlight on synaptic vesicles. *Trends Neurosci.* **13**, 83–87 (1990).
- Maycox, P. R., Deckwerth, T., Hell, J. W. & Jahn, R. Glutamate uptake by brain synaptic vesicles. Energy dependence of transport and functional reconstitution in proteoliposomes. *J. Biol. Chem.* **263**, 15423–15428 (1988).
- Hell, J. W., Maycox, P. R. & Jahn, R. Energy dependence and functional reconstitution of the gamma-aminobutyric acid carrier from synaptic vesicles. *J. Biol. Chem.* **265**, 2111–2117 (1990).
- Stern-Bach, Y., Russo, S., Neuman, M. & Rosenmund, C. A point mutation in the glutamate binding site blocks desensitization of AMPA receptors. *Neuron* **21**, 907–918 (1998).
- Raizen, D. M. & Avery, L. Electrical activity and behavior in the pharynx of *Caenorhabditis elegans*. *Neuron* **12**, 483–495 (1994).
- Lee, R. Y., Sawin, E. R., Chalfie, M., Horvitz, H. R. & Avery, L. EAT-4, a homolog of a mammalian sodium-dependent inorganic phosphate cotransporter, is necessary for glutamatergic neurotransmission in *Caenorhabditis elegans*. *J. Neurosci.* **19**, 159–167 (1999).
- Aihara, Y. *et al.* Molecular cloning of a novel brain-type Na⁺-dependent inorganic phosphate cotransporter. *J. Neurochem.* **74**, 2622–2625 (2000).
- Ashery, U., Betz, A., Xu, T., Brose, N. & Rettig, J. An efficient method for infection of adrenal chromaffin cells using the Semliki Forest virus gene expression system. *Eur. J. Cell Biol.* **78**, 525–532 (1999).
- Chen, C. & Okayama, H. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell Biol.* **7**, 2745–2752 (1987).

19. Hell, J. W., Maycox, P. R., Stadler, H. & Jahn, R. Uptake of GABA by rat brain synaptic vesicles isolated by a new procedure. *EMBO J.* **7**, 3023–3029 (1988).
20. Bekkers, J. M. & Stevens, C. F. Excitatory and inhibitory autaptic currents in isolated hippocampal neurons maintained in cell culture. *Proc. Natl Acad. Sci. USA* **88**, 7834–7838 (1991).
21. Rosenmund, C., Feltz, A. & Westbrook, G. L. Synaptic NMDA receptor channels have a low open probability. *J. Neurosci.* **15**, 2788–2795 (1995).
22. Hell, J. W. & Jahn, R. in *Cell Biology: a Laboratory Handbook* 1st edn (ed. Celis, J. E.) 567–574 (Academic, New York, 1994).
23. Lombard-Platet, G. & Jalilou, P. Funnel-well SDS-PAGE: a rapid technique for obtaining sufficient quantities of low-abundance proteins for internal sequence analysis. *Biotechniques* **15**, 668–670, 672 (1993).

Acknowledgements

We thank J. Rettig for discussions and the help in the viral infection technique; D. Pommereit and Y. Stern-Bach for providing the non-desensitizing AMPA-receptor vectors GluR1L497Y and GluR2QL504Y-IRES-DsRed; M. Druminski, D. Diezmann, A. Bührmann, I. Herfort and N. Narayanan for their technical assistance; P. Holroyd for critical reading of this manuscript. We also thank The HHMI Biopolymer/W.M. Keck Foundation, Biotechnology Resource Laboratory at Yale University for amino-acid sequencing.

Correspondence and requests for materials should be addressed to R.J. (e-mail: rjahn@gwdg.de).

Topological restriction of SNARE-dependent membrane fusion

Francesco Parlati, James A. McNew*, Ryouichi Fukuda*, Rebecca Miller, Thomas H. Söllner & James E. Rothman

Cellular Biochemistry and Biophysics Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, Box 251 New York, New York 10021, USA

To fuse transport vesicles with target membranes, proteins of the SNARE (soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors) complex must be located on both the vesicle (v-SNARE) and the target membrane (t-SNARE)¹. In yeast, four integral membrane proteins, Sed5, Bos1, Sec22 and Bet1 (refs 2–6), each probably contribute a single helix to form the SNARE complex that is needed for transport from endoplasmic reticulum to Golgi^{7–11}. This generates a four-helix bundle¹², which ultimately mediates the actual fusion event¹³. Here we explore how the anchoring arrangement of the four helices affects their ability to mediate fusion. We reconstituted two populations of phospholipid bilayer vesicles, with the individual SNARE proteins distributed in all possible combinations between them. Of the eight non-redundant permutations of four subunits distributed over two vesicle populations, only one results in membrane fusion. Fusion only occurs when the v-SNARE Bet1 is on one membrane and the syntaxin heavy chain Sed5 and its two light chains, Bos1 and Sec22, are on the other membrane where they form a functional t-SNARE. Thus, each SNARE protein is topologically restricted by design to function either as a v-SNARE or as part of a t-SNARE complex.

Full-length versions of Sed5, Bos1 and Bet1 were produced as glutathione-S-transferase (GST) fusion proteins in *Escherichia coli* (Fig. 1, lanes 1, 3 and 4). The full-length form of Sec22 was produced in *E. coli* and purified by nickel chromatography by an amino-terminal His₆ tag (Fig. 1, lane 2). To test whether these recombinant proteins, like their native counterparts, have an intrinsic ability to bind one another, we mixed purified Sec22, Bos1 and Bet1 with a bacterial cell lysate containing GST–Sed5, and isolated protein

complexes with glutathione–agarose beads. Almost stoichiometric amounts of Bos1, Sec22 and Bet1 were bound to GST–Sed5 (Fig. 1, lane 5; see legend).

We took an empirical approach, by testing all eight possible (non-redundant) distributions of four distinct SNAREs among two populations (Table 1), to determine which endoplasmic reticulum (ER)-to-Golgi SNARE(s) are required in the same or different vesicle populations to permit membrane fusion. Altogether, there are eight non-redundant combinations in which (1) all four SNAREs are in one vesicle population and none is in the other; (2) any three distinct SNAREs are in one population and the fourth distinct SNARE is in the other; and (3) any two SNAREs are in one population and the remaining two distinct SNAREs are in the other.

We tested all possible combinations of two SNAREs reconstituted in acceptor and two SNAREs in donor vesicles (Fig. 2). Equimolar amounts of SNAREs were mixed in buffer containing octylglucoside before adding lipids and preparing proteoliposomes as described¹³. All possible combinations of these donor and acceptor liposomes (including cases not shown in Table 1 where one SNARE was

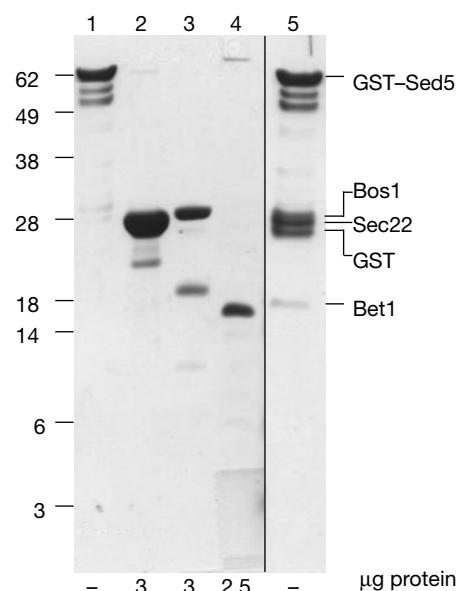


Figure 1 Sed5, Bos1, Sec22 and Bet1 form a quaternary complex. GST–Sed5 lysate was mixed with purified Sec22 (lane 2), Bos1 (lane 3) or Bet1 (lane 4), and 5% of input material was separated. The three SNAREs were isolated in the presence (lane 5) but not in the absence (data not shown) of GST–Sed5. GST–Sed5 isolated in the absence of Sec22, Bos1 and Bet1 is shown for comparison (lane 1). Proteins were analysed by SDS–PAGE and Coomassie blue staining. Note that Bet1 appears at substoichiometric levels because of its poor staining by Coomassie relative to other SNAREs. The stoichiometry of GST–Sed5/Bos1/Sec22/Bet1 in the SNARE complex is 1/1.2/1.2/0.6, as determined by densitometry. Because of the slight differences in the Bet1 migration in lanes 4 and 5 (caused by the presence of non-ionic detergent in lane 4), we confirmed the identity of Bet1 in the SNARE complex isolated in lane 5 by immunoblotting and N-terminal sequencing. Molecular masses in kilodaltons are indicated to the left.

Table 1 Non-redundant combinations of reconstituted SNAREs

Combinations	Acceptor liposomes				Donor liposomes	
4:0	Sed5	Bos1	Sec22	Bet1	–	
3:1	Sed5	Bos1	Sec22	Bet1	Bet1	
	Sed5		Sec22	Bet1	Bos1	
	Sed5	Bos1	Sec22	Bet1	Sec22	
2:2	Sed5	Bos1			Sed5	
	Sed5		Sec22		Sec22	Bet1
	Sed5			Bet1	Bos1	Sec22

* Present address: Department of Biochemistry and Cell Biology, Rice University, MS140, 6100 Main Street, PO Box 1892, Houston, Texas 77005, USA (J.A.M.); Department of Biotechnology, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan (R.F.).

replaced by a redundant one) were mixed, and the increase in 7-nitrobenz-2-oxa-1,3-diazole (NBD) fluorescence was monitored¹³ (Fig. 2b) and converted to rounds of fusion of donor vesicles^{13,14}. None of the combinations led to significant fusion.

We next tested all of the combinations in which acceptor liposomes contain three distinct SNAREs with donor liposomes containing a single SNARE, including all of the non-redundant combinations (Table 1) as well as other controls in which one SNARE was redundant (Fig. 3). Notably, only one combination,

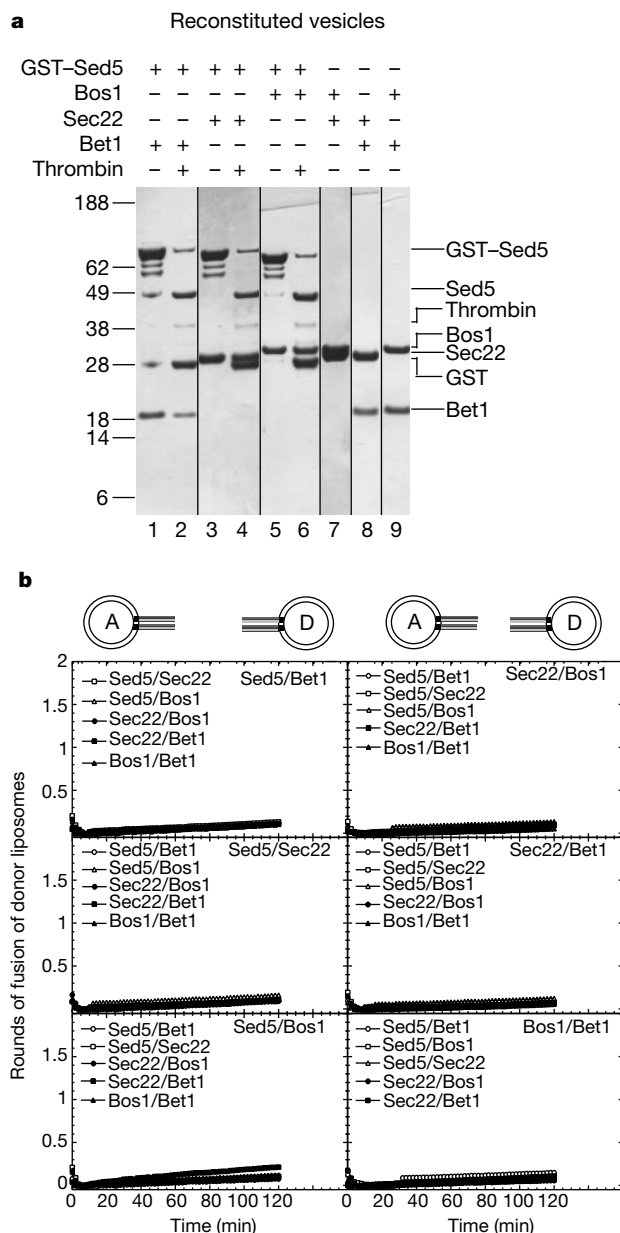


Figure 2 ER-Golgi SNAREs do not cause fusion when reconstituted into liposomes in a 2:2 fashion. **a**, The following sets of two ER-Golgi SNAREs were reconstituted into acceptor and donor liposomes and the protein recovery was monitored by SDS-PAGE and Coomassie blue staining: GST-Sed5/Bet1 (lane 1); GST-Sed5/Sec22 (lane 3); GST-Sed5/Bos1 (lane 5); Sec22/Bos1 (lane 7); Sec22/Bet1 (lane 8); and Bos1/Bet1 (lane 9). The GST tag was excised from liposomes containing GST-Sed5 (lanes 1, 3 and 5) by thrombin cleavage (lanes 2, 4 and 6). Molecular masses in kilodaltons are indicated to the left. **b**, Donor liposomes were mixed with acceptor liposomes at 37 °C as indicated, and the increase in fluorescence was monitored over 120 min. Fluorescence values were converted into a measure of rounds of fusion as described¹⁴.

Bet1 donor liposomes mixed with Sed5/Bos1/Sec22 acceptor liposomes, was fusogenic (Fig. 3c, top left panel).

We recovered roughly 33% of the added GST-Sed5/Bos1/Sec22 and 55% of the added Bet1 in acceptor and donor vesicles. After protease digestion, about 70% of this protein was orientated outwards (Fig. 3a, compare lanes 1 and 2), and all the reconstituted protein was insensitive to carbonate extraction (data not shown), confirming that the transmembrane domains were inserted into the lipid bilayer. Typical molar lipid to SNARE ratios for reconstituted donor and acceptor liposomes were about 60 and 200, respectively. Assuming a vesicle diameter of 50 nm and 65 Å per phospholipid^{13,14}, we estimate that there are about 380 copies of Bet1 per donor liposome and 115 copies of t-SNARE complex (Sed5/Bos1/Sec22) per acceptor liposome. As we add acceptor liposomes in molar excess (roughly 17-fold more lipid) over donor liposomes, there will be uncombined Bet1 after a single round of fusion, so that donor vesicles can theoretically undergo several rounds of fusion.

According to our calculation, Bet1 donor liposomes can theor-

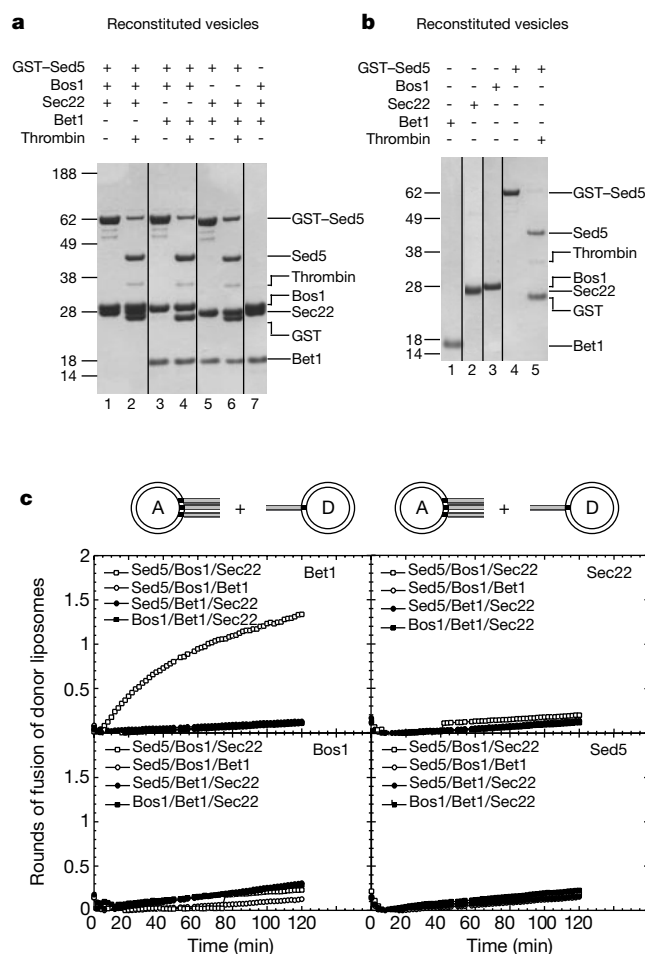


Figure 3 ER-Golgi SNAREs can fuse lipid membranes when reconstituted asymmetrically into liposomes. **a**, The following sets of three ER-Golgi SNAREs were reconstituted into acceptor liposomes: GST-Sed5/Bos1/Sec22 (lane 1); GST-Sed5/Bos1/Bet1 (lane 3); GST-Sed5/Sec22/Bet1 (lane 5); and Sec22/Bos1/Bet1 (lane 7). Protein recovery was monitored by SDS-PAGE and Coomassie blue staining. The GST tag was removed by thrombin cleavage (lanes 2, 4 and 6). Molecular mass in kilodaltons is indicated to the left. **b**, Bet1 (lane 1), Sec22 (lane 2), Bos1 (lane 3) and GST-Sed5 (lane 4) were reconstituted into donor liposomes. GST-Sed5 donor liposomes were further treated with thrombin to excise the GST tag (lane 5). **c**, Acceptor and donor liposomes were mixed as indicated at 37 °C, and fluorescence values measured over 2 h were converted into rounds of fusion¹⁴.

etically undergo 3.3 rounds of fusion. At the end of a 120-min period, Bet1 donor liposomes undergo 1.3–1.7 rounds of fusion (see Figs 3–5) or 40–50% of the theoretical maximum potential for fusion, underscoring the efficiency of SNARE-mediated lipid-bilayer fusion. The kinetics of fusion were similar to those observed for yeast vacuolar SNAREs¹⁵ and mammalian exocytic SNAREs¹³. In these cases, functional docking is the slowest step; fusion after SNARE complex formation is much more rapid. The slow kinetics of SNARE assembly have been shown in the case of exocytic SNAREs to be due to negative regulation by the N-terminal domain of syntaxin, which results in a ‘closed’ basal conformation of the t-SNARE¹⁴.

Bet1 donor liposomes did not fuse with acceptor liposomes in which Bet1 replaced any other SNARE (Fig. 3c, top left panel), consistent with the expectation that all four SNAREs are required to fuse liposomes. Sec22 donor liposomes do not fuse with Sed5/Bos1/Bet1 acceptor liposomes (Fig. 3c, top right panel). Bos1 donor liposomes do not fuse with Sed5/Sec22/Bet1 acceptor liposomes (Fig. 3c, bottom left panel), nor do Sed5 donor liposomes fuse with Sec22/Bos1/Bet1 acceptor liposomes (bottom right panel). Therefore, there is a strict topological requirement for Sed5, Bos1 and Sec22 in one lipid bilayer and Bet1 in another lipid bilayer. As expected, no fusion was observed when all four SNAREs were present in the same vesicle population (data not shown).

To confirm that all four SNAREs are required for fusion, we made acceptor liposome populations that lacked Bos1, Sec22 or Sed5 and then tested them for fusion with Bet1 donor liposomes (Fig. 4a). Liposomes lacking either Bos1, Sec22 or Sed5 did not fuse. We also

found that the Sed5/Bos1/Sec22 liposomes required the presence of Bet1 on the opposing membrane (Fig. 4b), eliminating the unlikely possibility that Sed5/Bos1/Sec22 liposomes are intrinsically fusogenic, and confirming the requirement for Bet1. Conversely, Bet1 donor liposomes do not fuse with protein-free acceptor liposomes (Fig. 4b), confirming the requirement for Sed5, Bos1 and Sec22 in acceptor liposomes.

As another control, we confirmed that fusion of Bet1 donor liposomes with Sed5/Bos1/Sec22 acceptor liposomes was completely inhibited by adding excess cytosolic domain of Bet1 (Fig. 4b). This ‘free’ SNARE effectively titrated out any uncombined Sed5/Bos1/Sec22 in acceptor liposomes, preventing the formation of *trans*-SNARE complexes with Bet1 donor liposomes. This result establishes the importance of SNARE complex formation in this fusion process.

The results that fusion between bilayers by isolated SNAREs is topologically restricted are consistent with the finding¹⁶, using a cell-free assay, that Bet1 is required on the transport vesicle membrane and Sed5 is required on the Golgi membrane. However, there was no requirement in this study for Bos1 or Sec22 at the Golgi membrane. This difference may somehow reflect the indirect methods used to reach this conclusion from cell extracts¹⁶, as distinct from our direct control of topology using pure proteins and lipids. The non-syntaxin SNAREs Bos1, Sec22 and Bet1 were all originally called v-SNAREs⁸ because they were all present in the ER and were heavily enriched in ER-derived coat protein II (COPII) transport vesicles^{9,17}. The assignment of Bos1 and Sec22 as light chains of a t-SNARE that can function at the Golgi, warranted by our detailed analysis of their function in fusion, is consistent with existing data indicating that Bos1, Sec22 and Sed5 are present in the Golgi, although they cycle rapidly through the ER and are thus dynamically localized^{16,18,19}.

After membrane fusion, all four SNAREs reside in a common membrane in the form of a ‘*cis*’-SNARE complex, which must then be disrupted to allow the constituent SNAREs to be used again²⁰. In

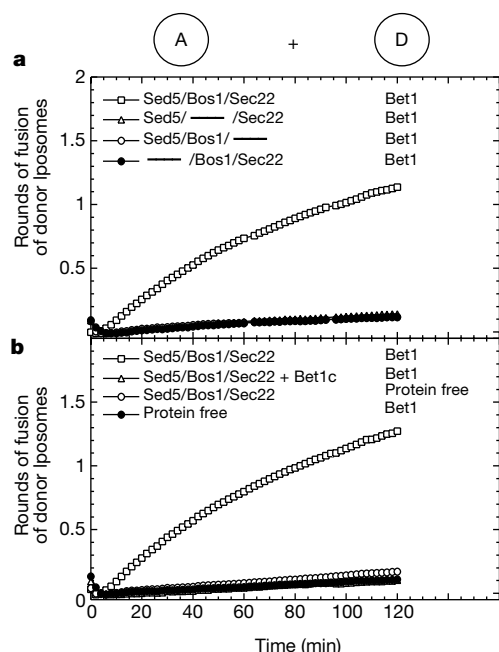


Figure 4 A trimeric t-SNARE and monomeric v-SNARE is required for membrane fusion. **a**, Sed5, Bos1 and Sec22 are obligate constituents of the ER–Golgi t-SNARE liposomes. Sed5/Bos1/Sec22 (open squares), Sed5/Sec22 (open triangles), Sed5/Bos1 (open circles) and Sec22/Bos1 (filled circles) were reconstituted into acceptor liposomes and mixed with Bet1-reconstituted donor liposomes at 37 °C. Fluorescence was monitored over a 2-h period and converted into rounds of fusion¹⁴. **b**, Fusion of acceptor and donor liposomes is v-SNARE- and t-SNARE-dependent and can be inhibited with the cytosolic domain of Bet1. Sed5/Bos1/Sec22 acceptor liposomes (open squares, open triangles and open circles) or protein-free acceptor liposomes (filled circles) were mixed with Bet1 donor liposomes (open squares, open triangles and filled circles) or protein-free donor liposomes (open circles) in the absence (open squares, open circles, filled circles) or presence (open triangles) of Bet1 cytosolic domain (Bet1c) at 37 °C. Fluorescence was monitored over a 2-h period and values were converted into rounds of fusion¹⁴.

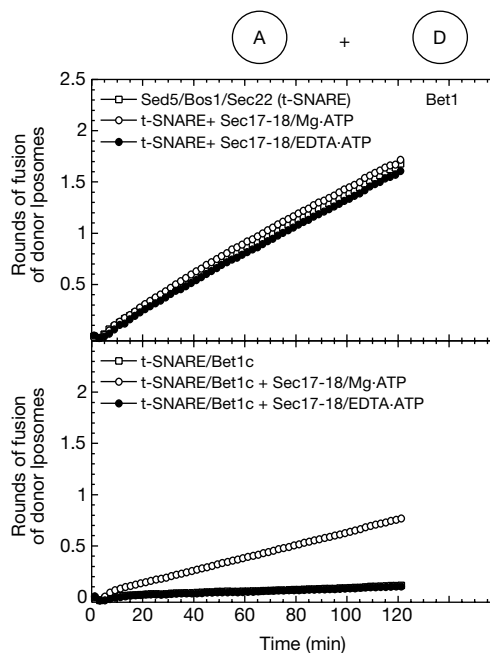


Figure 5 ER–Golgi v-SNARE/t-SNARE complex is a substrate for Sec17/Sec18. The t-SNAREs Sed5/Bos1/Sec22 (top panel) or v-SNARE/t-SNARE complex consisting of Sed5/Bos1/Sec22 and the cytosolic domain of Bet1 (bottom panel) were reconstituted into acceptor liposomes and mixed with Bet1 donor liposomes in the presence of buffer alone (open squares) or Sec17/Sec18 under ATP-hydrolysing (open circles) or ATP-non-hydrolysing (filled circles) conditions.

animals, α -SNAP (soluble *N*-ethylmaleimide-sensitive factor attachment protein) and NSF (*N*-ethylmaleimide-sensitive factor) mediate this process using ATP hydrolysis as a source of energy²¹. The yeast homologues of SNAP and NSF are Sec17 and Sec18, respectively.

When we mixed Sed5/Bos1/Sec22 acceptor liposomes with Bet1 donor liposomes, 1.7 rounds of fusion took place (Fig. 5, top panel). Neither the kinetics nor the extent of fusion were altered when Sec17/Sec18 under ATP-hydrolysing conditions were added to this fusion reaction (Fig. 5, top panel). We reconstituted a *cis*-SNARE complex consisting of Sed5, Bos1, Sec22 and the cytosolic domain of Bet1 into acceptor liposomes (Fig. 5, bottom panel). Owing to titration of the t-SNARE by the soluble fragment of the v-SNARE Bet1, we did not observe fusion when we mixed these proteoliposomes with donor liposomes containing full-length Bet1 (Fig. 5, bottom panel, open squares). We carried out the same fusion reaction in the presence of Sec17 and Sec18 with either Mg-ATP (open circles) or ATP in the absence of free Mg (with EDTA; filled circles). When Sec18 (which uses Mg-ATP as substrate) was allowed to function, the inhibition by the Bet1 cytosolic domain was significantly reduced, resulting in half of the rate of fusion that was observed when acceptor liposomes were prepared in the absence of the Bet1 cytosolic domain (compare open circles in top and bottom panels of Fig. 5).

One role of the ATPase NSF (Sec18) and its co-factor SNAP (Sec17) is to dissociate SNARE complexes; here we have confirmed that this is also the case for ER–Golgi SNARE complex. It follows that the *cis*-SNARE complex composed of Sed5, Bos1, Sec22 and Bet1 is a substrate for disruption by the NSF ATPase. Because it is not membrane-anchored, cytosolic Bet1 will dissociate from acceptor liposomes following SNARE complex disassembly, allowing Sed5/Bos1/Sec22 on acceptor liposomes to engage Bet1 on donor liposomes, resulting in the observed membrane fusion. Re-assembly of the freed cytosolic Bet1 onto acceptor liposomes will compete with this, explaining why fusion is only partially restored. Fusion can proceed in the presence of NSF and α -SNAP with Mg-ATP, which are present throughout the cytoplasm, because SNARE-Epins engaging in fusion are resistant to disruption at that moment²².

How might fusion be topologically restricted? The effects on fusion by substitution of various lipid anchors for natural protein anchors imply that the rod-like helical bundle of the SNARE complex exerts inward force on the membrane anchors, pulling them by taut linkers; lengthening the linker to either the v-SNARE or t-SNARE anchors inhibits fusion²³. Altogether, this suggests that the topological relationship of the linker to membrane anchor is a critical component in the force transduction mechanism that triggers fusion. In this context, switching the topology of bilayer insertion of one SNARE relative to the others would probably represent a big change. Such a change might slow the rate of SNARE assembly or reduce the stability of SNARE complexes if it introduces an extra energy barrier to either the transition state or the final state of the SNARE complex, respectively. At one extreme, the wrong topology might uncouple SNARE assembly from bilayer fusion, at the other it might prevent the SNARE complex from fully assembling. Preliminary experiments using the reconstituted ER–Golgi SNAREs indicate that *trans*-SNARE complexes do not form with non-fusogenic SNARE topologies (data not shown; assayed by vesicle binding²⁴).

Topological restriction may contribute to the compartmental specificity²⁵ of SNARE-dependent fusion; this layer of specificity would not have been apparent in studies involving only isolated cytoplasmic domains^{26,27}. Regulatory systems such as Rabs and tethers²⁸ may asymmetrically activate or position individual SNAREs in a fashion compatible with topological restriction. The principle of topological restriction of fusion requires that designated subunits of a SNARE complex must reside in opposite bilayers

for fusion to occur. This inherent polarity means that each SNARE has a genetically programmed role in a fusion reaction that restricts it to function as either a v-SNARE or a t-SNARE. Thus, we can deduce that Bos1 and Sec22 comprise the two light chains of the t-SNARE for this fusion reaction, whose syntaxin heavy chain is Sed5. Bos1 and Sec22 are both required in the same bilayer with the syntaxin Sed5 for fusion to occur and no other combination of non-syntaxin chains will substitute. Bet1 is the v-SNARE because it must reside in the opposite bilayer for fusion to occur. This functional identification of the t-SNARE for ER–Golgi transport is consistent with the cooperative binding of the light chains Bos1 and Sec22 to Sed5 suggesting that these proteins form a stable ternary complex¹¹. Consistent with topological restriction, gene deletion experiments show that fusion of native vacuoles will only occur if Nyv1 and Vam3 reside in opposite membranes²⁰. □

Methods

Plasmid construction

All plasmids were propagated in the *E. coli* strain DH5 α and standard DNA manipulation techniques were used. Plasmid pJM22 (ref. 29; pGEX-2T-based; Pharmacia) codes for full-length Sed5 with an N-terminal GST tag. We amplified the coding region for Bet1 and Bos1 by polymerase chain reaction (PCR) from *Saccharomyces cerevisiae* genomic DNA and inserted into pJM22 digested with *Nde*I and *Bam*HI to create pFP284 (GST–Bet1) and pFP285 (GST–Bos1). The Sec22 open reading frame was amplified by PCR from genomic DNA and cloned into pET-28a (Novagen) to create pFP266 (His₆–Sec22). The cytosolic domain of Bet1 (amino acids 1–124) was amplified by PCR from genomic DNA and cloned into pET15b digested with *Nde*I and *Bam*HI creating pFP306 (His₆Bet1 Δ TMD). Plasmid pFP307 codes for an N-terminal His₆-tagged Sec17 and was generated from a PCR product coding for the Sec17 coding region (292 amino acids) and inserted into the *Bam*HI and *Hind*III sites of pQE-9 (Qiagen). Plasmid pFP308 codes for an N-terminal His₆-tagged Sec18 and was generated from a PCR product encompassing the Sec18 coding region (2,814 coding nucleotides and 227 non-coding nucleotides downstream from the stop codon) and inserted into the *Bam*HI and *Hind*III sites of pQE-9 (Qiagen).

Protein expression and purification

Plasmids used for protein expression were transformed into *E. coli* strain BL21 (DE3). Transformed cells were grown at 37 °C to an absorbance at 600 nm of 0.8 when protein expression was induced with 1 mM IPTG (Boehringer Mannheim) for 4–6 h at 37 °C for GST–Bet1 (pFP284), GST–Bos1 (pFP285), His₆–Bet1 Δ TMD (pFP306), His₆–Sec17 (pFP307) and His₆–Sec18 (pFP308). For GST–Sed5 (pJM22) and His₆–Sec22 (pFP266), we grew cells at 37 °C to an absorbance at 600 nm of 0.4, and transferred them to 16 °C until an absorbance at 600 nm of 0.8 was reached. Protein expression was induced with 1 mM IPTG (Boehringer-Mannheim) for at least 20 h at 16 °C. We collected cells expressing GST-tagged proteins by centrifugation and lysed them by one passage through an Avestin (Ottawa, Ontario) cell disrupter at > 10,000 p.s.i. in buffer A (25 mM Tris pH 8.0, 400 mM KCl, 10% glycerol, 4% Triton X-100, 5 mM β -mercaptoethanol, 1 mM phenylmethylsulphonyl fluoride (PMSF)) and lysates were clarified by centrifugation.

Lysates containing GST-tagged proteins were bound to glutathione-agarose for 1 h and packed beads were washed with buffer B (25 mM Tris pH 8.0, 400 mM KCl, 10% glycerol, 1% β -octylglucoside, 2 mM β -mercaptoethanol, 1 mM PMSF). GST–Sed5 was eluted with buffer B containing 10 mM glutathione. Bet1 was eluted by incubating beads with buffer B containing 0.015 units per μ l thrombin (2 h at room temperature), whereas Bos1 was eluted by incubating beads with buffer B (100 mM instead of 400 mM KCl) containing 0.05 units per μ l thrombin (2 h at 37 °C) and then adjusted to 400 mM KCl. Cells expressing His₆–Sec22 were also lysed in buffer A, and the clarified lysate was bound to Ni-NTA agarose and subsequently washed with buffer D (25 mM HEPES pH 7.2, 500 mM KCl, 10% glycerol, 1% β -octylglucoside, 2 mM β -mercaptoethanol, 50 mM imidazole). Proteins were then eluted with a 50 mM to 1 M imidazole gradient (in buffer D). We purified His₆–Sec17 and His₆–Sec18 as described³⁰. We purified His₆–Bet1 Δ TMD from inclusion bodies as described²⁹.

Reconstitution into liposomes and thrombin cleavage

We reconstituted GST–Sed5, Bos1, Sec22 and Bet1 into acceptor or donor liposomes essentially as described^{13,14} with the following modifications. (1) Each SNARE was mixed to a final concentration of 27 μ M in total volume of 500 μ l (volume adjusted with buffer B) and added to acceptor lipid film (100 μ l for donor lipid film). (2) NycoDenz gradients were run with 400 mM KCl instead of 100 mM KCl. After liposome recovery, liposomes containing GST–Sed5 were treated with 0.05 units per μ l thrombin (Sigma) for 2 h at room temperature and subsequently thrombin was inactivated with 2 mM AEBSF (Calbiochem). For Bet1 Δ TMD inhibition studies, we added 1 nmol of His₆–Bet1 Δ TMD to the mixture of acceptor liposomes 30 min before the addition of donor liposomes.

Fusion assays

Fusion experiments were conducted essentially as described^{13,14,23}. Briefly, we mixed

acceptor liposomes (45 μ l) and donor liposomes (5 μ l) in a 96-well FluoroNunc microtitre plate (Nunc). In some cases a Sec17/Sec18 cocktail was added to the fusion reactions consisting of 15 μ g ml⁻¹ Sec17, 87 μ g ml⁻¹ Sec18, 0.97 mg ml⁻¹ ATP, 9.7 mM creatine phosphate (Boehringer Mannheim), 0.1 mg ml⁻¹ creatine phosphate kinase (Boehringer Mannheim), 1.3 mM Mg₂Cl₂ (or 1.3 mM EDTA). Microtitre plates were then placed in a Fluoroscan II (Labsystems) equilibrated at 37 °C and NBD fluorescence was followed. Conversion to a measure of rounds of fusion was done as described¹⁴.

Received 17 April; accepted 27 July 2000.

- Söllner, T. *et al.* SNAP receptors implicated in vesicle targeting and fusion. *Nature* **362**, 318–324 (1993).
- Novick, P., Field, C. & Schekman, R. Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway. *Cell* **21**, 205–215 (1980).
- Newman, A. P. & Ferro-Novick, S. Characterization of new mutants in the early part of the yeast secretory pathway isolated by a [3H]mannose suicide selection. *J. Cell Biol.* **105**, 1587–1594 (1987).
- Shim, J., Newman, A. P. & Ferro-Novick, S. The BOS1 gene encodes an essential 27-kD putative membrane protein that is required for vesicular transport from the ER to the Golgi complex in yeast. *J. Cell Biol.* **113**, 55–64 (1991).
- Dascher, C., Ossig, R., Gallwitz, D. & Schmitt, H. D. Identification and structure of four yeast genes (SLY) that are able to suppress the functional loss of YPT1, a member of the RAS superfamily. *Mol. Cell Biol.* **11**, 872–885 (1991).
- Hardwick, K. G. & Pelham, H. R. SED5 encodes a 39-kD integral membrane protein required for vesicular transport between the ER and the Golgi complex. *J. Cell Biol.* **119**, 513–521 (1992).
- Newman, A. P., Shim, J. & Ferro-Novick, S. BET1, BOS1, and SEC22 are members of a group of interacting yeast genes required for transport from the endoplasmic reticulum to the Golgi complex. *Mol. Cell Biol.* **10**, 3405–3414 (1990).
- Sogaard, M. *et al.* A rab protein is required for the assembly of SNARE complexes in the docking of transport vesicles. *Cell* **78**, 937–948 (1994).
- Lian, J. P., Stone, S., Jiang, Y., Lyons, P. & Ferro-Novick, S. Ypt1p implicated in v-SNARE activation. *Nature* **372**, 698–701 (1994).
- Stone, S. *et al.* Bet1p activates the v-SNARE Bos1p. *Mol. Biol. Cell* **8**, 1175–1181 (1997).
- Sacher, M., Stone, S. & Ferro-Novick, S. The syntaxin-like region of Sed5p. *J. Biol. Chem.* **272**, 17134–17138 (1997).
- Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* **395**, 347–353 (1998).
- Weber, T. *et al.* SNAREpins: minimal machinery for membrane fusion. *Cell* **92**, 759–772 (1998).
- Parlati, F. *et al.* Rapid and efficient fusion of phospholipid vesicles by the alpha-helical core of a SNARE complex in the absence of an N-terminal regulatory domain. *Proc. Natl Acad. Sci. USA* **96**, 12565–12570 (1999).
- Fukuda, R. *et al.* Functional architecture of an intracellular t-SNARE. *Nature* **407**, 198–202 (2000).
- Cao, X. & Barlowe, C. Asymmetric requirements for a Rab GTPase and SNARE proteins in fusion of COPII vesicles with acceptor membranes. *J. Cell Biol.* **149**, 55–66 (2000).
- Rexach, M. F., Latterich, M. & Schekman, R. W. Characteristics of endoplasmic reticulum-derived transport vesicles. *J. Cell Biol.* **126**, 1133–1148 (1994).
- Boehm, J., Ulrich, H. D., Ossig, R. & Schmitt, H. D. Kex2-dependent invertase secretion as a tool to study the targeting of transmembrane proteins which are involved in ER→Golgi transport in yeast. *EMBO J.* **13**, 3696–3710 (1994).
- Barrowman, J., Sacher, M. & Ferro-Novick, S. TRAPP stably associates with the Golgi and is required for vesicle docking. *EMBO J.* **19**, 862–869 (2000).
- Nichols, B. J., Ungermann, C., Pelham, H. R., Wickner, W. T. & Haas, A. Homotypic vacuolar fusion mediated by t- and v-SNAREs. *Nature* **387**, 199–202 (1997).
- Söllner, T., Bennett, M. K., Whiteheart, S. W., Scheller, R. H. & Rothman, J. E. A protein assembly–disassembly pathway *in vitro* that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. *Cell* **75**, 409–418 (1993).
- Weber, T. *et al.* SNAREpins are functionally resistant to disruption by N-ethylmaleimide-sensitive fusion protein (NSF) and α -soluble NSF attachment protein (SNAP). *J. Cell Biol.* **149** (2000).
- McNew, J. A., Weber, T., Engelman, D. M., Söllner, T. H. & Rothman, J. E. The length of the flexible SNAREpin juxtamembrane region is a critical determinant of SNARE-dependent fusion. *Mol. Cell* **4**, 415–421 (1999).
- McNew, J. A. *et al.* Close is not enough: SNARE-dependent membrane fusion requires an active mechanism that transduces force to membrane anchors. *J. Cell Biol.* **150**, 105–118 (2000).
- McNew, J. A. *et al.* Compartmental specificity of cellular membrane fusion encoded in SNARE proteins. *Nature* **407**, 153–159 (2000).
- Fasshauer, D., Antonin, W., Margittai, M., Pabst, S. & Jahn, R. Mixed and non-cognate SNARE complexes. Characterization of assembly and biophysical properties. *J. Biol. Chem.* **274**, 15440–15446 (1999).
- Yang, B. *et al.* SNARE interactions are not selective. Implications for membrane fusion specificity. *J. Biol. Chem.* **274**, 5649–5653 (1999).
- Mellman, I. & Warren, G. The road taken: past and future foundations of membrane traffic. *Cell* **100**, 99–112 (2000).
- McNew, J. A. *et al.* Gos1p, a *Saccharomyces cerevisiae* SNARE protein involved in Golgi transport. *FEBS Lett.* **435**, 89–95 (1998).
- Whiteheart, S. W. *et al.* N-ethylmaleimide-sensitive fusion protein: a trimeric ATPase whose hydrolysis of ATP is required for membrane fusion. *J. Cell Biol.* **126**, 945–954 (1994).

Acknowledgements

We would like to thank R. J. Johnston for technical support. Research was supported by an NIH grant (to J.E.R.) and postdoctoral fellowships of the Medical Research Council of Canada (F.P.), the NIH (J.M.) and the Japanese Society for the Promotion of Science (R.F.).

Correspondence and requests for materials should be addressed to J.E.R. (e-mail: j-rothman@ski.mskcc.org).

Functional architecture of an intracellular membrane t-SNARE

Ryouchi Fukuda*, James A. McNew*, Thomas Weber*, Francesco Parlati, Thomas Engel*, Walter Nickel*, James E. Rothman & Thomas H. Söllner

Cellular Biochemistry & Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, Box 519, New York, New York 10021, USA

* Present addresses: Department of Biotechnology, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan (R.F.); Department of Biochemistry and Cell Biology, Rice University, MS140, 6100 Main Street, PO Box 1892, Houston, Texas 77251-1892, USA (J.A.M.); Institute for Gene Therapy and Molecular Medicine, Box 1496, Mount Sinai School of Medicine, 1425 Madison Avenue, New York, New York 10029-6574, USA (T.W.); Institut für Arterioskleroseforschung an der Westfälischen Wilhelms-Universität, Domagkstraße 3, D-48149 Münster, Germany (T.E.); Biochemie-Zentrum Heidelberg, Ruprecht-Karls-Universität, Im Neuenheimer Feld 328, D-69120 Heidelberg, Germany (W.N.).

Lipid bilayer fusion is mediated by SNAREs (soluble N-ethylmaleimide-sensitive factor attachment protein receptors) located on the vesicle membrane (v-SNAREs) and the target membrane (t-SNAREs)^{1,2}. The assembled v-SNARE/t-SNARE complex consists of a bundle of four helices, of which one is supplied by the v-SNARE and the other three by the t-SNARE³. For t-SNAREs on the plasma membrane, the protein syntaxin⁴ supplies one helix and a SNAP-25 protein⁵ contributes the other two. Although there are numerous homologues of syntaxin on intracellular membranes⁶, there are only two SNAP-25-related proteins in yeast, Sec9 and Spo20, both of which are localized to the plasma membrane and function in secretion⁷ and sporulation⁸, respectively. What replaces SNAP-25 in t-SNAREs of intracellular membranes? Here we show that an intracellular t-SNARE is built from a 'heavy chain' homologous to syntaxin and two separate non-syntaxin 'light chains'. SNAP-25 may thus be the exception rather than the rule, having been derived from genes that encoded separate light chains that fused during evolution to produce a single gene encoding one protein with two helices.

To test our model of the functional architecture of the t-SNAREs on intracellular membranes (Fig. 1), we studied the SNAREs involved in the well characterized process of fusion of yeast vacuoles^{9–12}. Homotypic fusion of yeast vacuoles is the final step in the inheritance of vacuoles from a mother cell to a daughter cell after bud formation in *Saccharomyces cerevisiae*. The SNARE proteins Vam3, Vam7, Vti1, Nyv1 and Ykt6 have been identified

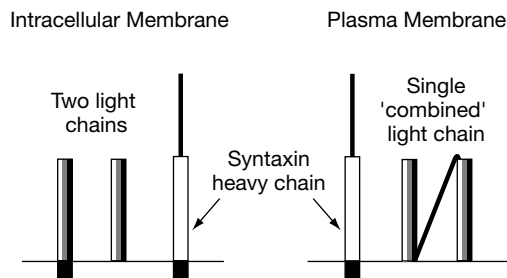


Figure 1 The model for the architecture of intracellular t-SNAREs. The plasma membrane t-SNARE is composed of one syntaxin heavy chain, which contributes one α -helix, and one combined light chain, SNAP-25, which contributes two α -helices. In comparison, most intracellular t-SNAREs are composed of one syntaxin heavy chain and two associated light chains. In the case of vacuolar homotypic fusion, Vam3 functions as a syntaxin heavy chain, and Vti1 and soluble Vam7 act as two light chains. In contrast to Vam7, most light chains are probably integral membrane proteins.

as being involved in this fusion process. Vam3 is a syntaxin family member; the others are non-syntaxin SNAREs. Vti1 has been found to be present in cells in a complex with the syntaxin Vam3, and also in distinct complexes with all but three of the other yeast syntaxins^{6,13}. Although Vti1 was originally classified as v-SNARE¹⁴, from the perspective outlined above it could just as easily be a shared light chain of many intracellular t-SNAREs, including a vacuolar t-SNARE. The function of Ykt6 in vacuolar fusion is not well defined, and this protein has been implicated in a variety of transport steps^{6,13}; it has also been proposed to be a fifth member of a single vacuolar SNARE complex¹¹, along with Vam3, Vam7, Vti1 and Nyv1. If this proposal is correct, then the vacuolar SNARE complex would have to be based on a different packing principle than the exocytic/neuronal SNARE complex¹¹.

We expressed recombinant Vam3, Vam7, Vti1, Ykt6 and Nyv1 as His₆- or glutathione-S-transferase (GST)-tagged proteins in *Escherichia coli*. Various combinations of purified SNAREs were incubated together to identify direct SNARE–SNARE interactions and higher order complexes. All binding studies were done under high stringency conditions in high salt (≥ 500 mM KCl) using full-length proteins in detergent solution. We analysed bound products by SDS–polyacrylamide gel electrophoresis (PAGE) and Coomassie blue staining (Fig. 2).

Immobilized GST–Vam3 (the syntaxin) efficiently binds Vti1 (Fig. 2a, lane 3) but not Vam7 (Fig. 2a, lane 2), which were added in a 5- and 25-fold excess, respectively. However, Vam7 was efficiently bound to the Vam3–Vti1 complex in roughly the same molar amount as Vti1 (Fig. 2a, lane 4). Binding of Vti1 and Vam7 to the

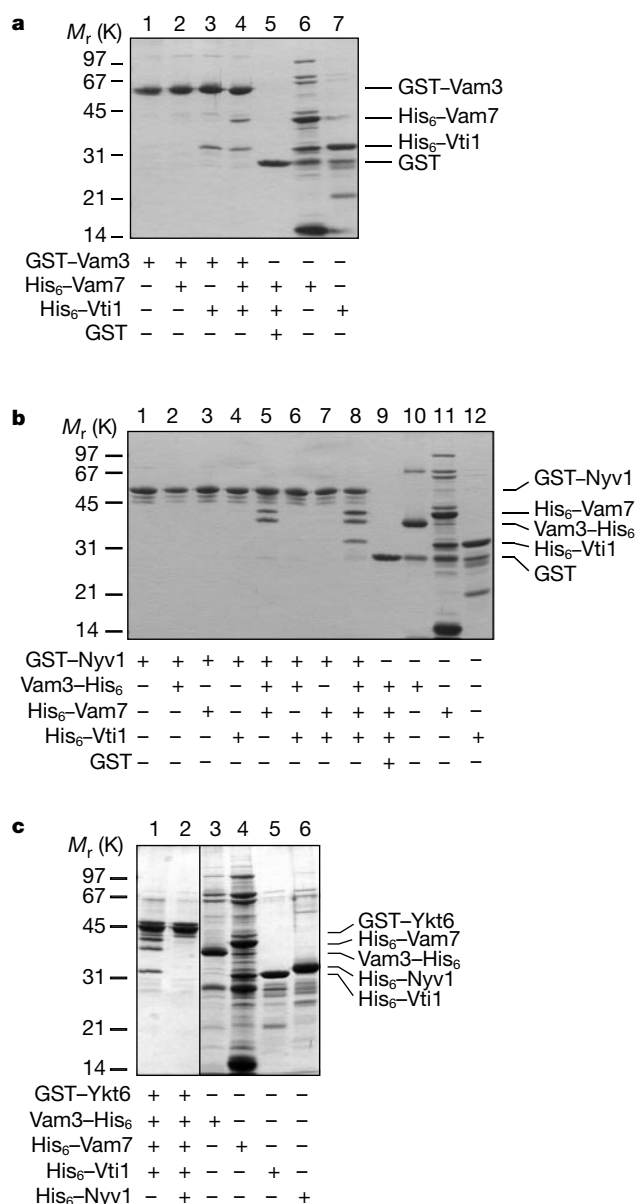


Figure 2 Vacuolar v-SNARE/t-SNARE complex. **a**, t-SNARE complex formation. Purified GST–Vam3 (12 μ g), or GST (5.5 μ g) immobilized on glutathione–agarose was incubated with purified His₆–Vam7 (207 μ g) and/or His₆–Vti1 (29 μ g) in the presence of 1% Triton X-100 in a final volume of 100 μ l. After overnight incubation at 4 °C, the beads were washed. One eighth of the reactions was analysed by SDS–PAGE, and proteins were stained with Coomassie Blue R 250. Lanes 6 and 7 correspond to 10% of input proteins. **b**, v-SNARE/t-SNARE complex formation. Purified GST–Nyv1 (11.2 μ g) or GST (5.5 μ g)

immobilized on glutathione–agarose were incubated with purified Vam3–His₆ (32 μ g), His₆–Vam7 (207 μ g) and/or His₆–Vti1 (29 μ g) as described in **a**. Lanes 10–12 correspond to 10% of input proteins. **c**, SNARE complex formation with Ykt6. Purified GST–Ykt6 (10 μ g) immobilized on glutathione–agarose was incubated with purified Vam3–His₆ (98 μ g), His₆–Vam7 (621 μ g) and His₆–Vti1 (86 μ g) in the presence or absence of His₆–Nyv1 (86 μ g), in a final volume of 200 μ l. Lanes 3–6 correspond to 0.33% of input proteins.

syntaphin Vam3 was stoichiometric (data not shown) and specific in that neither Vam7 nor Vti1 interacted with GST, and protein impurities in the SNARE preparations did not bind to GST or GST–Vam3 (Fig. 2a, lanes 5 and 6–7, respectively).

These data suggest that the syntaphin (Vam3)-containing vacuolar t-SNARE is a ternary complex consisting of a syntaphin heavy chain (Vam3) and two non-syntaphin light chains (Vti1 and Vam7), along the lines we proposed. The t-SNARE would then be expected to bind of the v-SNARE Nyv1 (Fig. 2b), and indeed, immobilized GST–Nyv1 efficiently binds the ternary t-SNARE complex (Fig. 2b, lane 8). Binding of the t-SNARE complex to the v-SNARE Nyv1 was saturable and stoichiometric (data not shown). Immobilized GST–Nyv1 does not bind efficiently to any of the individual t-SNARE subunits Vam3 (heavy chain), or Vam7 or Vti1 (the light chains) (Fig. 2b, lanes 2–4). Nor does Nyv1 bind the binary complex of Vam3–Vti1 (Fig. 2b, lane 6), which is stable (Fig. 2a, lane 3). However, Nyv1 does stabilize a complex of Vam3 and Vam7 (Fig. 2b, lane 5), which does not form appreciably in the absence of Nyv1 (Fig. 2a, lane 2), clearly illustrating the cooperativity of the interactions among the SNAREs.

As Ykt6 has been suggested to be a fifth component of the vacuolar SNARE complex¹¹, we determined whether Ykt6 binds the above tetrameric SNARE complex of Vam3–Vam7–Vti1–Nyv1. GST–Ykt6 binds the ternary t-SNARE (Vam3–Vam7–Vti1) (Fig.

2c, lane 1), but binding is blocked by equimolar amounts of Nyv1 (Fig. 2c, lane 2). In other words, Nyv1 and Ykt6 cannot bind to the t-SNARE complex at the same time. This directly establishes that the five vacuolar SNAREs can assemble into either of two alternative quaternary complexes, but not into a pentameric complex. On the basis of inhibition studies using antibodies against Ykt6 (ref. 11), Ykt6 appears to be involved in vacuolar fusion; and co-immunoprecipitation experiments from lysates showed that a pentameric SNARE complex containing all five SNAREs—Ykt6, Nyv1, Vam3, Vam7 and Vti1—can be isolated from membranes¹¹. Our results are not necessarily inconsistent with these data¹¹, as other unidentified proteins might be present in the immunoprecipitates characterized by western blotting using antibodies to known SNARE proteins. Alternatively, oligomeric complexes containing both of the tetrameric complexes that we identified might have been precipitated.

To test the functional capacity of the quaternary vacuolar SNARE complex, we reconstituted its v- and t-SNARE portions into lipid bilayer vesicles, by using established procedures². Recombinant vacuolar t-SNARE subunits (the syntaphin Vam3 and the light chains Vti1 and Vam7) were reconstituted into 'acceptor' liposomes, whereas the v-SNARE Nyv1 was reconstituted into fluorescent 'donor' liposomes² (Fig. 3a). Typical molar lipid-to-SNARE ratios for the reconstituted donor (v-SNARE) and acceptor (t-SNARE) liposomes were about 60 and 470, respectively. Like

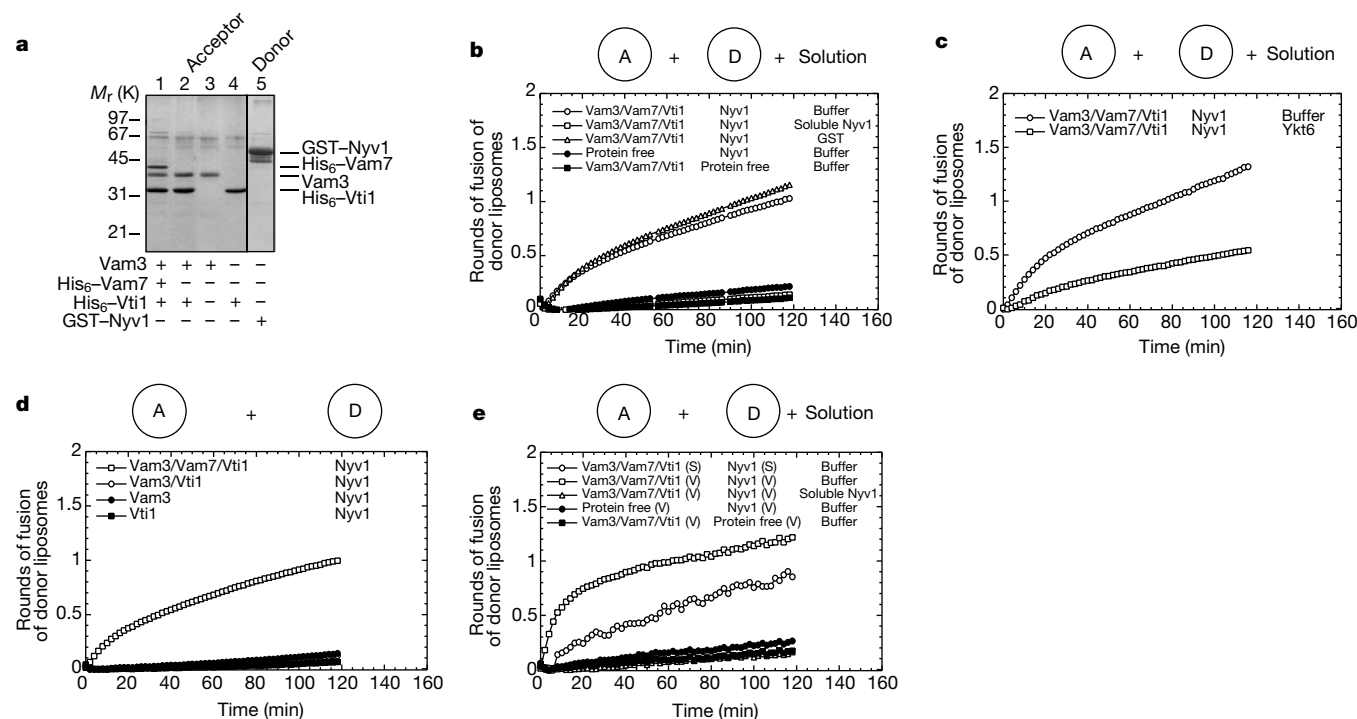


Figure 3 Fusion between vesicles containing vacuolar v- and t-SNAREs. **a**, Acceptor vesicles containing t-SNAREs and donor vesicles containing v-SNAREs. Protein patterns of purified acceptor liposomes containing various t-SNARE subunits and of donor liposomes containing the v-SNARE GST–Nyv1 were analysed by SDS–PAGE and Coomassie Blue R 250. Lanes 1–4 correspond to ~10% of acceptor vesicles; lane 5 to 100% of donor vesicles used in a typical fusion reaction. **b**, Donor vesicles (5 μ l; 4.3 nmol phospholipid) containing the v-SNARE and acceptor vesicles (45 μ l; 107 nmol phospholipid) containing t-SNAREs were mixed with 15 μ l of either reconstitution buffer, or GST–cytoplasmic domain of Nyv1 (soluble Nyv1; 2.86 mg ml^{–1}), corresponding to a fourfold molar excess over t-SNAREs or GST (1.48 mg ml^{–1}). Fusion of protein free vesicles was also tested with either donor v-liposomes or with acceptor t-liposomes. The temperature was shifted to 37 °C and NBD fluorescence monitored at 538 nm. After 2 h at 37 °C, *n*-dodecylmaltoside was added to solubilize liposomes and to de-quench the fluorescence of NBD. To normalize the experiments, the lowest NBD fluorescence signals were set to 0% and the highest signals, after addition of detergent, were set to 100%.

Rounds of fusion were determined using a standard curve¹⁶. **c**, Donor vesicles (5 μ l; 3.9 nmol phospholipid) containing the v-SNARE and 45 μ l of acceptor vesicles (109 nmol phospholipid) containing t-SNAREs were mixed with 20 μ l of either reconstitution buffer or GST–Ykt6 (0.62 mg ml^{–1}), corresponding to a 3.7-fold molar excess over v-SNAREs. Fusion was monitored as in **b**. **d**, Donor vesicles (5 μ l; 4.3 nmol phospholipid) containing the v-SNARE Nyv1 were mixed with acceptor vesicles (45 μ l; 73 nmol phospholipid) containing either the ternary t-SNARE complex or Vam3 and Vti1, or Vam3 or Vti1. Fusion was monitored as in **b**. **e**, Donor vesicles and acceptor vesicles were prepared with either standard lipid mixture (S) or vacuolar lipid mixture (V). To compare directly the fusion reactions and to correct for the different reconstitution efficiencies, donor liposomes with standard lipid mixture contained half of the amount of GST–Nyv1 used in **a–d**. Donor vesicles (5 μ l; 2.0 nmol phospholipid) containing the v-SNARE and acceptor vesicles (45 μ l 97 nmol phospholipid) containing t-SNAREs were mixed with 15 μ l of either reconstitution buffer, or GST–cytoplasmic domain of Nyv1, and fusion was monitored as **b**, except that reactions were incubated at 30 °C.

exocytic/neuronal SNAREs², the v-SNAREs reconstitute more efficiently than the t-SNAREs, resulting in a higher number of copies of v-SNAREs per donor liposome than t-SNAREs per acceptor liposome. Assuming a vesicle diameter of 50 nm and 65 Å² per phospholipid, we estimate that there are about 400 copies of the v-SNARE Nyv1 protein per donor vesicle and 50 copies of the ternary t-SNARE complex (consisting of Vam3, Vti1 and Vam7) in the acceptor vesicles. About 80% of v-SNAREs and 70% of t-SNAREs are exposed on the outside of the liposomes (data not shown). Fusion was measured by the increase in 7-nitrobenz-2-oxa-1,3-diazole (NBD) fluorescence¹⁵ and converted to rounds of fusion of donor liposome bilayers¹⁶.

Donor and acceptor liposomes formed of the standard lipid mixture (see Methods) fused efficiently at 37 °C (Fig. 3b, open circles). The v-SNARE-containing donor liposomes do not fuse with protein-free acceptor liposomes and vice versa, protein-free donor liposomes do not fuse with t-SNARE-containing acceptor liposomes (Fig. 3b, filled circles and filled squares, respectively). Furthermore, free cytoplasmic domain of Nyv1 completely abolished fusion (Fig. 3b, open squares), indicating that fusion required the formation of SNARE complexes between donor and acceptor liposomes. This 'soluble' v-SNARE titrates the t-SNAREs of acceptor vesicles, preventing them from engaging v-SNAREs on donor vesicles. Addition of Ykt6 in amounts equimolar with the t-SNARE complex (a 3.7-fold molar excess compared with the v-SNARE) also inhibited the fusion reaction by 60% (Fig. 3c). This result is consistent with our protein binding studies (Fig. 2c) and confirms that Ykt6 and Nyv1 compete for binding to the t-SNARE complex. In addition, Ykt6-containing liposomes (in which Ykt6 is anchored by a close mimic of its natural lipid modification) do not fuse with vacuolar t-SNARE-containing liposomes¹⁷.

To confirm that all four SNAREs are needed for fusion, we prepared acceptor liposomes in which one or several components of the t-SNARE were omitted during reconstitution. The different t-SNARE liposomes contain similar protein levels to allow direct comparisons of fusion efficiencies (Fig. 3a). Acceptor liposomes containing only the syntaxin Vam3, or only the light chain Vti1, or both Vam3 and Vti1, did not fuse with the Nyv1 donor liposomes (Fig. 3d). Fusion required the ternary t-SNARE complex.

The kinetics of the above fusion reactions are similar to those observed for the mammalian exocytic SNAREs². The speed of fusion was significantly increased when a mixture of lipids representative of vacuoles (see Methods) was used and when the incubation was carried out at 30 °C, a common temperature for yeast growth (Fig. 3e). Now, the half-time for the first round of fusion was reduced to about 7 min, which is comparable to or faster than the kinetics of fusion of isolated native vacuolar membranes¹⁸.

Our observations support a general model for intracellular t-SNAREs (Fig. 1) in which a syntaxin family member and two associated light chains form a ternary complex, as first suggested from sequence analysis¹⁹, which would be expected to consist of a three-helix bundle³. The t-SNARE complex that we report for the yeast vacuole is stable in dilute solution, but this need not always be the case as long as the complex assembles when bound to the v-SNARE in bilayers.

The non-syntaxin SNARE Vti1 forms complexes with most of the yeast syntaxins in intracellular compartments⁶. As for the vacuolar t-SNARE, Vti1 may well be a common light chain of other intracellular t-SNAREs, rather than a promiscuous v-SNARE²⁰. In general terms, the recognition that non-syntaxin SNAREs can function either as subunits of t-SNAREs or as v-SNAREs provides a straight forward basis for grouping what has become a long and sometimes confusing list of SNARE proteins into a simpler pattern of functional units. Moreover, by no longer equating syntaxins with t-SNAREs and non-syntaxin SNAREs with v-SNAREs, several apparent difficulties vanish. For example, the obvious and unresolved question—why would cells need so many v-SNAREs in

relation to t-SNAREs—is no longer pertinent. A non-syntaxin SNARE common to the t-SNAREs of different compartments is perfectly consistent with the SNARE hypothesis, but would contradict the hypothesis if the same SNARE could function as a v-SNARE. Finally, combinatorial use of the same syntaxin with different sets of light chains can generate greater diversity in traffic patterns than would be predicted from the number of syntaxins alone, which may be specially important in the Golgi and possibly in endosomes. According to analysis of the yeast genome⁶, there are only five syntaxins in the intracellular compartments of yeast (Vam3, Sed5, Ufe1, Pep12 and Tlg2), which, however, could generate a variety of t-SNARE complexes in combination with the 14 yeast v-SNAREs. (On the basis of the new classification¹⁹, Tlg1 is a member of the TSL (t-SNARE-like protein) family and not a direct syntaxin homologue (see also <http://itsa.ucsf.edu/~mostov/supplementary>). Indeed, the Golgi-localized syntaxin5 (Sed5 in yeast), which appears to form at least two distinct sets of complexes with non-syntaxin SNAREs (that is, light chains)^{6,21}, is one example for such a combinatorial code.

Although pairing of SNAREs between vacuoles is important for tight docking and therefore indirectly important for fusion, it has been proposed that SNARE pairing is insufficient for fusion. Instead, a still unknown principle that is somehow activated by SNARE pairing has been invoked to explain the fusion of bilayers²². However, our results with isolated vacuolar SNARE proteins directly establish that pairing of these SNARE proteins in itself results in spontaneous and efficient bilayer fusion. Essentially, fusion of vacuoles was largely resistant to excess *N*-ethylmaleimide-sensitive factor (NSF), which is believed to disrupt all SNARE complexes²². But it is now known that, whereas *cis*-SNARE complexes are disrupted by NSF, under the same conditions the *trans*-SNARE complexes (SNAREpins) engaged in fusion are resistant to NSF and soluble NSF attachment proteins (SNAPs)²³.

Calcium flux and phosphorylation/dephosphorylation reactions are important for fusion of yeast vacuoles^{18,24}. Yet, fusion by the isolated vacuolar SNARE proteins clearly does not depend on additional proteins or enzyme systems, and it is not affected by the addition or absence of calcium ions (data not shown). We propose that the well documented requirements for a phosphatase and calcium flux for fusion of native vacuoles may reflect regulatory mechanisms that are superimposed upon the basic and general core fusion machinery, SNARE proteins. As vacuolar fusion is temporally and spatially regulated to occur during M to G1 phase and within the bud²⁵, such regulatory mechanisms of one kind or another are obviously expected. Indeed, it is well established that phosphorylation can prevent the assembly of t-SNAREs in other cases^{26,27}. Therefore, phosphorylation may also negatively regulate vacuolar SNAREs, and dephosphorylation may relieve such inhibitory effects on membrane fusion. □

Methods

Plasmids and protein purification

The coding sequences of NYV1, VTI1, VAM7 and VAM3 were amplified by PCR from *S. cerevisiae* genomic DNA (Novagen) and cloned into either pET15b, pET28a (Novagen) or pGEX4T-3 (Pharmacia) expression vectors. YKT6 was subcloned from pMS130 (ref. 28) into pGEX4T-3 (Pharmacia). We expressed all SNARE proteins in *E. coli* and purified them by either glutathione- or nickel-affinity chromatography. More details are available from the authors on request.

Protein-binding assay

For preparation of GST–Nyv1, GST–Vam3, GST–Ykt6, or GST affinity matrices, we incubated lysates prepared from cells containing pGEX–Nyv1, pGEX–Vam3, pGEX–Ykt6 or pGEX–4T-3 at 4 °C for 2 h with glutathione agarose equilibrated in binding buffer A (25 mM HEPES–KOH pH 7.4, 500 mM KCl, 10% glycerol, 1 mM DTT and 1% Triton X-100). After centrifugation at 1,000 g for 5 min, the beads were washed 6 times with binding buffer A. SNARE proteins and bovine serum albumin (500 µg ml⁻¹) were added and the beads were incubated at 4 °C over night. The beads were then washed three times with 400 µl of binding buffer A, and 30 µl of binding buffer B (25 mM HEPES–KOH pH 7.4, 100 mM KCl, 10% glycerol, 1 mM DTT, 1% Triton X-100) was added. One eighth

of each sample was mixed with 4 × Laemmli SDS–PAGE buffer, boiled for 4 min at 98 °C and resolved by SDS–PAGE.

Protein reconstitution into vesicles

For acceptor vesicles containing t-SNAREs, 60 µl of Vam3 (2.9 mg ml⁻¹), 200 µl of His₆–Vam7 (8.2 mg ml⁻¹) and/or 20 µl of His₆–Vti1 (14 mg ml⁻¹) were mixed and 20 µl of 10% octyl-β-D-glucopyranoside and 250 µl of reconstitution buffer were added. After incubation at 4 °C for 20 h, 500 µl of the reaction was used for reconstitution of acceptor vesicles. For donor vesicles containing the v-SNARE, 100 µl of GST–Nyy1 (3.6 mg ml⁻¹) was used. SNAREs were reconstituted as described² except that all the buffers used for reconstitution contained 500 mM KCl. The principal lipid components in the liposomes were 85% DOPC (1,2-dioleoyl phosphatidylcholine) and 15% DOPS (1,2-dioleoyl phosphatidylserine). Typical lipid concentration in the recovered Nycodenz fractions containing the vesicles were 0.7–0.9 mM for donor vesicles and 1.8–2.5 mM for acceptor vesicles. Concentrations of GST–Nyy1 in donor vesicle fractions were 0.7–0.8 mg ml⁻¹ and those of t-SNARE complex in acceptor vesicles were 0.45–0.5 mg ml⁻¹, corresponding to a 16% and 9% reconstitution efficiency. For preparation of vesicles containing vacuolar lipids^{29,30}, we used the following lipid mixture: DOPC (1,2-dioleoyl phosphatidylcholine)/DOPE (1,2-dioleoyl phosphatidylethanolamine)/soybean PI (phosphatidylinositol)/DOPS (1,2-dioleoyl phosphatidylserine)/DOPA (1,2-dioleoyl phosphatidic acid)/heart cardiolipin in a 53:20:19:5:2:1 molar ratio. All lipids were obtained from Avanti Polar Lipids. The reconstitution efficiencies for v-SNAREs and t-SNAREs in vacuolar lipid mixtures were 8% and 9%, respectively.

Lipid mixing assay

Lipid mixing assay was carried out as described^{2,15}, except that *n*-dodecylmaltoside was used to determine the maximum NBD signal instead of Triton X-100. Rounds of fusion were determined by using a calibration curve which was prepared as described¹⁶, except that GST–Nyy1 was used instead of VAMP to reconstitute vesicles and that the mixtures of labelled and unlabelled lipids corresponded to one-, two-, three-, four-, or sixfold dilutions of donor fluorescent lipid. The equation used to convert the percentage dodecylmaltoside signal to rounds of fusion is $y = (0.52247 \times e^{(0.032993x)}) - (0.527 \times e^{(-0.074915x)})$ where y is the rounds of fusion and x is the percentage dodecylmaltoside signal at a given time interval.

Received 17 April; accepted 27 July 2000.

- Söllner, T. *et al.* SNAP receptors implicated in vesicle targeting and fusion. *Nature* **362**, 318–324 (1993).
- Weber, T. *et al.* SNAREpins: minimal machinery for membrane fusion. *Cell* **92**, 759–772 (1998).
- Sutton, R. B., Fasshauer, D., Jahn, R. & Bringer, A. T. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* **395**, 347–353 (1998).
- Bennett, M. K., Calakos, N. & Scheller, R. H. Syntaxin: a synaptic protein implicated in docking of synaptic vesicles at presynaptic active zones. *Science* **257**, 255–259 (1992).
- Oyler, G. A. *et al.* The identification of a novel synaptosomal-associated protein, SNAP-25, differentially expressed by neuronal subpopulations. *J. Cell Biol.* **109**, 3039–3052 (1989).
- Pelham, H. R. SNAREs and the secretory pathway—lessons from yeast. *Exp. Cell Res.* **247**, 1–8 (1999).
- Rossi, G., Salminen, A., Rice, L. M., Bringer, A. T. & Brennwald, P. Analysis of a yeast SNARE complex reveals remarkable similarity to the neuronal SNARE complex and a novel function for the C terminus of the SNAP-25 homolog, Sec9. *J. Biol. Chem.* **272**, 16610–16617 (1997).
- Neiman, A. M. Prosopore membrane formation defines a developmentally regulated branch of the secretory pathway in yeast. *J. Cell Biol.* **140**, 29–37 (1998).
- Nichols, B. J., Ungermann, C., Pelham, H. R., Wickner, W. T. & Haas, A. Homotypic vacuolar fusion mediated by t- and v-SNAREs. *Nature* **387**, 199–202 (1997).
- Ungermann, C. & Wickner, W. Vam7p, a vacuolar SNAP-25 homolog, is required for SNARE complex integrity and vacuole docking and fusion. *EMBO J.* **17**, 3269–3276 (1998).
- Ungermann, C. *et al.* Three v-SNAREs and two t-SNAREs, present in a pentameric cis-SNARE complex on isolated vacuoles, are essential for homotypic fusion. *J. Cell Biol.* **145**, 1435–1442 (1999).
- Sato, T. K., Darow, T. & Emr, S. D. Vam7p, a SNAP-25-like molecule, and Vam3p, a syntaxin homolog, function together in yeast vacuolar protein trafficking. *Mol. Cell Biol.* **18**, 5308–5319 (1998).
- Nichols, B. J. & Pelham, H. R. SNAREs and membrane fusion in the Golgi apparatus. *Biochim. Biophys. Acta* **1404**, 9–31 (1998).
- Fischer von Mollard, G., Nothwehr, S. F. & Stevens, T. H. The yeast v-SNARE Vti1p mediates two vesicle transport pathways through interactions with the t-SNAREs Sed5p and Pep12p. *J. Cell Biol.* **137**, 1511–1524 (1997).
- Struck, D. K., Hoekstra, D. & Pagano, R. E. Use of resonance energy transfer to monitor membrane fusion. *Biochemistry* **20**, 4093–4099 (1981).
- Parlati, F. *et al.* Rapid and efficient fusion of phospholipid vesicles by the alpha-helical core of a SNARE complex in the absence of an N-terminal regulatory domain. *Proc. Natl Acad. Sci. USA* **96**, 12565–12570 (1999).
- McNew, J. A. *et al.* Compartmental specificity of cellular membrane fusion encoded in SNARE proteins. *Nature* **407**, 153–159 (2000).
- Peters, C. & Mayer, A. Ca²⁺/calmodulin signals the completion of docking and triggers a late step of vacuole fusion. *Nature* **396**, 575–580 (1998).
- Weimbs, T., Mostov, K., Low, S. H. & Hofmann, K. A model for structural similarity between different SNARE complexes based on sequence relationships. *Trends Cell Biol.* **8**, 260–262 (1998).
- Fischer von Mollard, G. & Stevens, T. H. The *Saccharomyces cerevisiae* v-SNARE Vti1p is required for multiple membrane transport pathways to the vacuole. *Mol. Biol. Cell* **10**, 1719–1732 (1999).
- Hay, J. C., Chao, D. S., Kuo, C. S. & Scheller, R. H. Protein interactions regulating vesicle transport between the endoplasmic reticulum and Golgi apparatus in mammalian cells. *Cell* **89**, 149–158 (1997).
- Ungermann, C., Sato, K. & Wickner, W. Defining the functions of trans-SNARE pairs. *Nature* **396**, 543–548 (1998).
- Weber, T. *et al.* SNAREpins are functionally resistant to disruption by NSF and alpha-SNAP. *J. Cell Biol.* (in the press).
- Peters, C. *et al.* Control of the terminal step of intracellular membrane fusion by protein phosphatase 1. *Science* **285**, 1084–1087 (1999).

- Weisman, L. S., Bacallao, R. & Wickner, W. Multiple methods of visualizing the yeast vacuole permit evaluation of its morphology and inheritance during the cell cycle. *J. Cell Biol.* **105**, 1539–1547 (1987).
- Foster, L. J. *et al.* Binary interactions of the SNARE proteins syntaxin-4, SNAP23, and VAMP-2 and their regulation by phosphorylation. *Biochemistry* **37**, 11089–11096 (1998).
- Risinger, C. & Bennett, M. K. Differential phosphorylation of syntaxin and synaptosome-associated protein of 25 kDa (SNAP-25) isoforms. *J. Neurochem.* **72**, 614–624 (1999).
- McNew, J. A. *et al.* Ykt6p, a prenylated SNARE essential for endoplasmic reticulum–Golgi transport. *J. Biol. Chem.* **272**, 17776–17783 (1997).
- Schneider, R. *et al.* Electrospray ionization tandem mass spectrometry (ESI-MS/MS) analysis of the lipid molecular species composition of yeast subcellular membranes reveals acyl chain-based sorting/remodeling of distinct molecular species en route to the plasma membrane. *J. Cell Biol.* **146**, 741–754 (1999).
- Zinser, E. & Daum, G. Isolation and biochemical characterization of organelles from the yeast, *Saccharomyces cerevisiae*. *Yeast* **11**, 493–536 (1995).

Acknowledgements

We wish to thank B. Brügger for preparation of the vacuolar lipid mixture and T. Wolfe for help with the manuscript. This work was supported by a NIH grant (to J.E.R.) and postdoctoral fellowships of the Japan Society for the Promotion of Science (to R.F.), the Human Frontiers Science Program Organization (to W.N.), the Deutsche Forschungsgemeinschaft (to T.E.), the Medical Research Council of Canada (to F.P.), the National Institutes of Health (to J.A.M.), the Swiss National Science Foundation (to T.W.) and the European Molecular Biology Organization (to T.W.).

Correspondence and requests for materials should be addressed to T.S. (e-mail: t-sollner@ski.mskcc.org).

CAP defines a second signalling pathway required for insulin-stimulated glucose transport

Christian A. Baumann*, Vered Ribon*, Makoto Kanzaki†, Debbie C. Thurmond†, Silvia Mora†, Satoshi Shigematsu†, Perry E. Bickel‡, Jeffrey E. Pessin† & Alan R. Saltiel*

* Department of Physiology, University of Michigan School of Medicine, Ann Arbor, Michigan 48109, USA and Department of Cell Biology, Parke-Davis Pharmaceutical Research Division, Warner-Lambert Company, Ann Arbor, Michigan 48105, USA

† Department of Physiology & Biophysics, The University of Iowa, Iowa City, Iowa 52242, USA

‡ Department of Internal Medicine and Department of Cell Biology and Physiology, Washington University School of Medicine, St. Louis, Missouri 63110, USA

Insulin stimulates the transport of glucose into fat and muscle cells. Although the precise molecular mechanisms involved in this process remain uncertain, insulin initiates its actions by binding to its tyrosine kinase receptor, leading to the phosphorylation of intracellular substrates. One such substrate is the Cbl proto-oncogene product¹. Cbl is recruited to the insulin receptor by interaction with the adapter protein CAP, through one of three adjacent SH3 domains in the carboxy terminus of CAP². Upon phosphorylation of Cbl, the CAP–Cbl complex dissociates from the insulin receptor and moves to a caveolin-enriched, triton-insoluble membrane fraction³. Here, to identify a molecular mechanism underlying this subcellular redistribution, we screened a yeast two-hybrid library using the amino-terminal region of CAP and identified the caveolar protein flotillin. Flotillin forms a ternary complex with CAP and Cbl, directing the localization of the CAP–Cbl complex to a lipid raft subdomain of the plasma membrane. Expression of the N-terminal domain of CAP in 3T3-L1 adipocytes blocks the stimulation of glucose transport by insulin, without affecting signalling events that depend on phosphatidylinositol-3-OH kinase. Thus, localization of the Cbl–CAP complex to lipid rafts generates a pathway that is crucial in the regulation of glucose uptake.

To identify proteins that may participate in the localization and function of Cbl and CAP, we screened a yeast two-hybrid library derived from 3T3-L1 adipocytes⁴ using as bait the N-terminal region of CAP, which contains the sorbin homology domains but lacks the SH3 domains (CAP Δ SH3, Fig. 1a). Three independent clones were isolated encoding the protein flotillin⁵. To examine the binding specificity of flotillin and CAP, we incubated glutathione-S-sepharose-bound flotillin (GST–flotillin) with lysates from human embryonic kidney (HEK293T) cells overexpressing either Flag-epitope-tagged full-length CAP or Flag-epitope-tagged CAP Δ SH3. As seen in Fig. 1b, GST–flotillin bound to both Flag–CAP and Flag–CAP Δ SH3. It also precipitated endogenous Cbl from the extracts in the presence of CAP but not when CAP Δ SH3 was co-expressed. These data indicate that the interaction of flotillin with Cbl requires full-length CAP to form a flotillin–CAP–Cbl ternary complex.

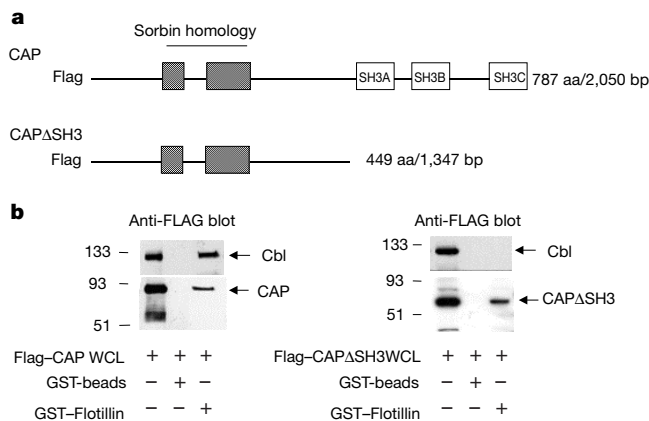


Figure 1 CAP and Cbl form a ternary complex with GST–flotillin. **a**, Flag-epitope-tagged expression constructs of CAP and CAP Δ SH3 mutant. **b**, Flag-epitope-tagged CAP (left) or CAP Δ SH3 (right) overexpression lysates from HEK293T cells were incubated with either glutathione agarose or glutathione-agarose-bound GST–flotillin for 1 h at 4 °C. Precipitates were subjected to SDS–polyacrylamide gel electrophoresis (PAGE), transferred to nitrocellulose and blotted with monoclonal Flag M2 antibody (Stratagene) and polyclonal c-Cbl antibody (Santa Cruz).

Triton-insoluble microdomains of the plasma membrane have been implicated in signal transduction^{6,7}. Flotillin is an integral membrane protein that associates with the triton-insoluble fraction enriched in the caveolar protein caveolin⁵. Moreover, flotillin can physically associate in hetero-oligomeric structures with caveolin in detergent-solubilized cells⁸. The physicochemical properties of proteins that are enriched in lipid rafts make it difficult to perform standard, triton-soluble, co-immunoprecipitation assays. We therefore examined the subcellular localization of CAP, CAP Δ SH3, endogenous caveolin and flotillin by confocal immunofluorescence microscopy of intact cells and by immunofluorescence following preparation of plasma membrane sheets. After transfection of 3T3-L1 adipocytes by electroporation⁹, both Flag–CAP and Flag–CAP Δ SH3 showed diffuse staining of the cytoplasm, with a significant portion found nonuniformly distributed around the plasma membrane (see Supplementary Information Fig. 1). Immunostaining of endogenous caveolin followed by a Texas Red secondary antibody revealed caveolin in the plasma membrane, as reported^{6,7}. Merging the two images shows that a significant portion of both CAP and CAP Δ SH3 proteins are found at the plasma membrane, similar to the endogenous caveolin I.

To examine the plasma membrane localization of caveolin in relation to flotillin, we prepared plasma membrane sheets from 3T3-L1 adipocytes and immunostained them with antibodies to caveolin and flotillin. Caveolin and flotillin were localized together in circular rosette structures in the plasma membrane (see Supplementary Information Fig. 2). The co-distribution of flotillin and caveolin in the plasma membrane was unaffected by insulin stimulation. These flotillin and caveolin-enriched structures probably consist of collections of caveolae or lipid rafts¹⁰.

To investigate further the dynamics of endogenous Cbl protein association with flotillin, we took a similar approach using immunofluorescence of isolated plasma membrane sheets (Fig. 2). As described above, exposure of cells to insulin for 2–60 min had no effect on the overall plasma membrane distribution of flotillin. Although Cbl was not detected in the plasma membrane sheets in the basal state, there was transient translocation of Cbl to the plasma membrane, which was apparent after 2 and 5 min of insulin stimulation. This time-dependent translocation of Cbl to the plasma membrane is similar to the time course of insulin-stimulated Cbl

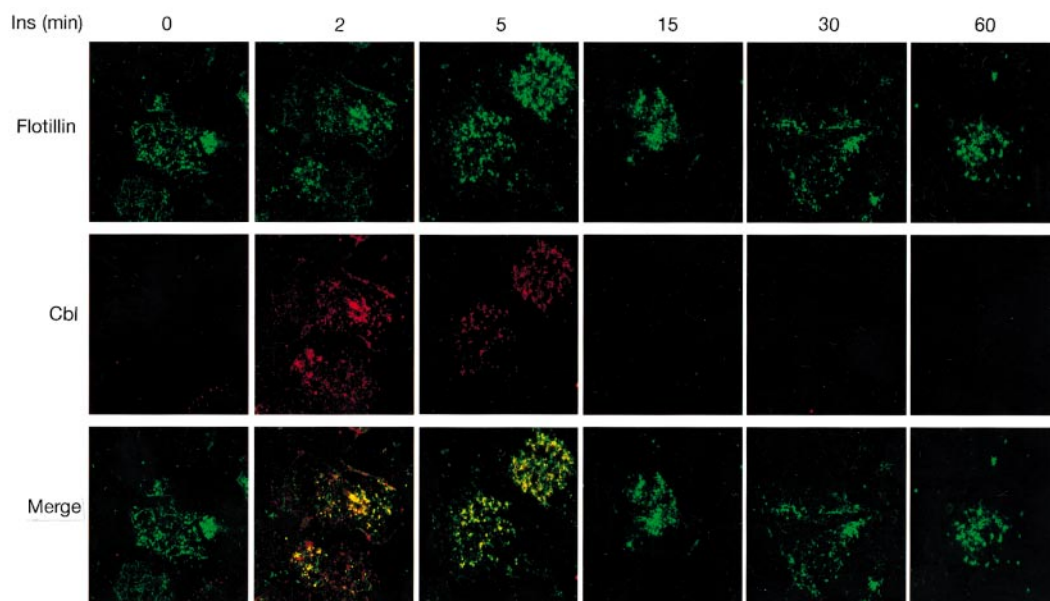


Figure 2 Translocation of Cbl to caveolin- and flotillin-containing plasma membrane subdomains of 3T3-L1 adipocytes. Representative confocal immunofluorescence images of flotillin and Cbl localized to plasma membrane sheets isolated from basal cells or those

treated with insulin (Ins) for the indicated times. The colocalization of FITC-labelled flotillin and Texas Red-labelled Cbl is shown.

tyrosine phosphorylation¹¹. Furthermore, merging of these images showed that Cbl localized with flotillin in rosette-type structures, again suggesting the insulin-dependent formation of a ternary flotillin–CAP–Cbl complex.

Finally, to assess biochemically the relationship between caveolin, flotillin and Cbl, we subjected basal and insulin-stimulated 3T3-L1 adipocytes to non-detergent homogenization and sucrose gradient fractionation (see Supplementary Information Fig. 3). Consistent with the immunofluorescence of isolated plasma membrane sheets, caveolin was found primarily in the low-density membrane regions of these gradients (fractions 3 and 4), and was not affected by insulin treatment. Similarly, flotillin was found in the fractions containing caveolin. Cbl was primarily confined to the dense regions of the gradient (fractions 8–12) in the basal state. However, following insulin stimulation, a portion of the Cbl protein was redistributed to the low-density fractions containing caveolin and flotillin. Together, these data provide strong evidence that caveolin and flotillin are colocalized in distinct subdomains of the plasma membrane, and that insulin induces the translocation of Cbl to these caveolin- and flotillin-enriched compartments through its interaction with CAP.

To assess the potential role of CAP in early signalling events, we next overexpressed CAP or CAP Δ SH3 in 3T3-L1 adipocytes and evaluated insulin-stimulated tyrosine phosphorylation. Both proteins were found in the triton-soluble and insoluble fractions (Fig. 3a), and expressed at levels around fivefold higher than the endogenous CAP protein (data not shown). Expression of CAP or CAP Δ SH3 had no effect on the tyrosine phosphorylation of the insulin receptor β -subunit or its predominant substrate, IRS-1 (Fig. 3b). Similarly, there was no change in the ability of insulin to stimulate the phosphorylation of two downstream targets, AKT/protein kinase B and mitogen-activated protein kinase (MAPK), that are activated as a consequence of well characterized phosphorylation cascades (Fig. 3b). As AKT activation occurs as a direct consequence of phosphatidylinositol-3-OH kinase (PI(3)K) stimulation¹², these data also show that the expression of CAP Δ SH3 did not affect the stimulation of PI(3)K activity by insulin.

The phosphorylation of Cbl by insulin requires the recruitment of

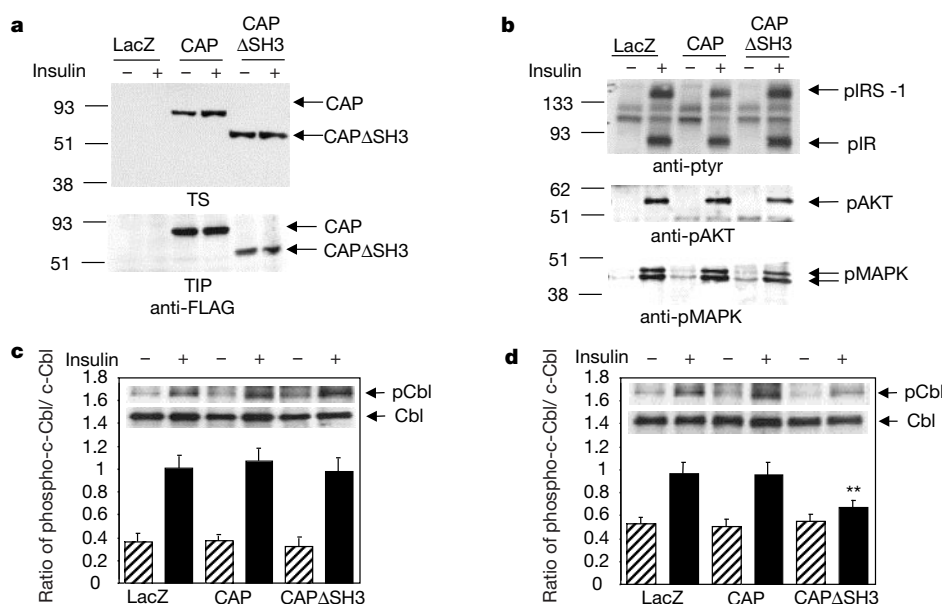


Figure 3 CAP Δ SH3 mutant blocks the translocation of phosphorylated Cbl into lipid rafts. **a**, A representative FLAG immunoblot of the triton-soluble (TS) fraction (top) and triton-insoluble pellet (TIP) fraction (bottom) of 3T3-L1 adipocytes expressing LacZ, CAP or CAP Δ SH3. **b**, Lysates from basal and insulin-treated adipocytes expressing LacZ, CAP or CAP Δ SH3 were immunoblotted with anti-phosphotyrosine (top), phospho-AKT (middle) and phospho-MAPK (bottom) antibodies. **c**, Cbl was immunoprecipitated from the triton-soluble fraction of lysates from basal and insulin-treated adipocytes expressing LacZ, CAP

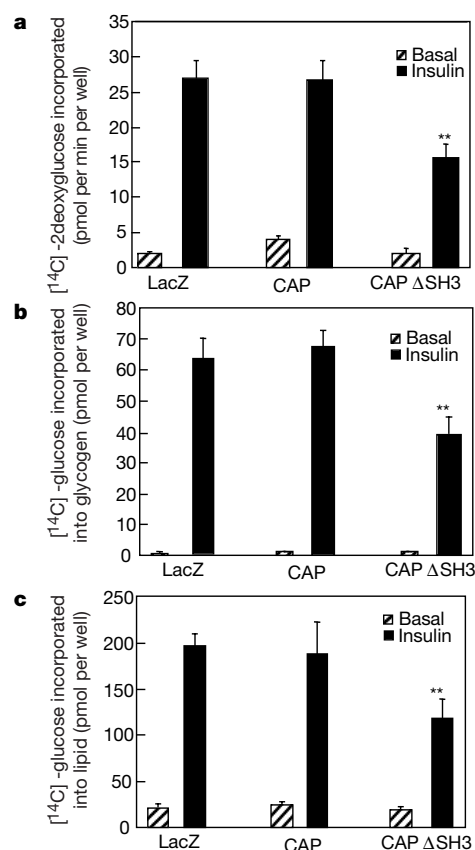


Figure 4 Overexpression of CAP Δ SH3, but not CAP, attenuates the metabolic effects of insulin. 3T3-L1 adipocytes were transfected by electroporation and the stimulation of [14 C]-2-deoxyglucose uptake (**a**), [14 C]-glucose incorporation into glycogen (**b**) and [14 C]-glucose incorporation into lipid (**c**) by insulin were measured. Results are the mean $\pm$ standard error of three separate experiments, each performed in triplicate. Double asterisk: significant difference, $P < 0.01$, $n = 3$.

or CAP Δ SH3. Immunoprecipitates were resolved by SDS–PAGE, immunoblotted with monoclonal phosphotyrosine antibodies and re-probed with the Cbl antibody. A representative blot is shown within the graph. Films were scanned and quantified, and the ratio of total Cbl to phospho-Cbl was calculated. **d**, Cbl was immunoprecipitated from the triton-insoluble pellet fraction of lysates from basal and insulin-treated adipocytes expressing LacZ, CAP or CAP Δ SH3. Immunoprecipitates were resolved and analysed as described in Fig. 2c. Double asterisk: significant difference, $P < 0.01$, $n = 4$.

the protein to the insulin receptor by CAP². As CAP Δ SH3 does not bind either Cbl or the insulin receptor, overexpression of CAP or CAP Δ SH3 had no effect on insulin-stimulated Cbl phosphorylation in the triton-soluble fraction (Fig. 3c). However, CAP Δ SH3 overexpression significantly reduced the appearance of tyrosine phosphorylated Cbl in the triton-insoluble fraction (Fig. 3d). These data indicate that the expression of the N-terminal flotillin-binding domain of CAP may be able to block the interaction of the endogenous CAP–Cbl complexes with flotillin, thus preventing the appearance of phosphorylated Cbl in the triton-insoluble plasma membrane subdomain.

To investigate the role of the association of the CAP–Cbl complex with flotillin in insulin action, we evaluated the effect of CAP Δ SH3 expression on insulin-stimulated [¹⁴C]2-deoxyglucose uptake in 3T3-L1 adipocytes (Fig. 4a). Although wild-type CAP had no effect, expression of CAP Δ SH3 led to a reduction of about 50% in insulin-stimulated glucose uptake, without affecting basal activity. Similar experiments were done to evaluate the effect of LacZ, CAP or CAP Δ SH3 on glycogen and lipid synthesis (Fig. 4b, c). As observed for glucose uptake, the expression of CAP Δ SH3 inhibited the insulin-stimulated incorporation of [¹⁴C]glucose into glycogen and lipid, whereas expression of CAP had no effect. Considering that the average efficiency of transfection by electroporation in these cells is about 50% (ref. 9), the observed reduction in glucose utilization may indicate that the expression of the CAP Δ SH3 mutant almost completely blocked the metabolic actions of insulin.

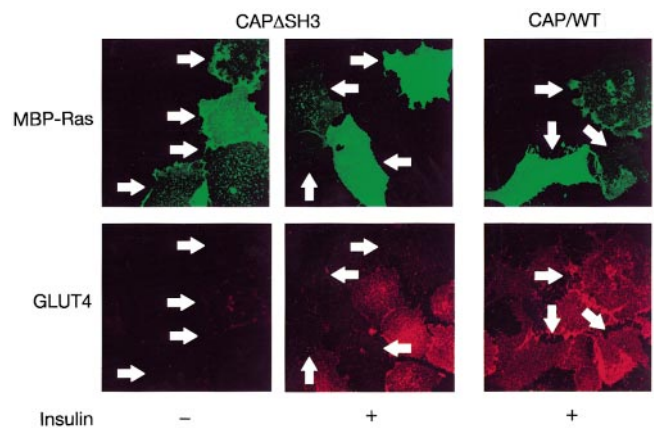


Figure 6 CAP Δ SH3 inhibits insulin-stimulated endogenous GLUT4 but not GLUT1 translocation. Differentiated 3T3-L1 adipocyte nuclei were microinjected with 0.2 mg ml⁻¹ of MBP-Ras plus either the CAP or CAP Δ SH3 cDNA. After 24 h recovery, cells were treated without or with 100 nM insulin for 30 min. Plasma membrane sheets were prepared and MBP-Ras was detected with a FITC-labelled anti-MBP antibody and GLUT4 with a Cy5-labelled anti-GLUT4 antibody. These are representative images obtained from five independent experiments and visualization of 60–91 individual plasma membrane sheets.

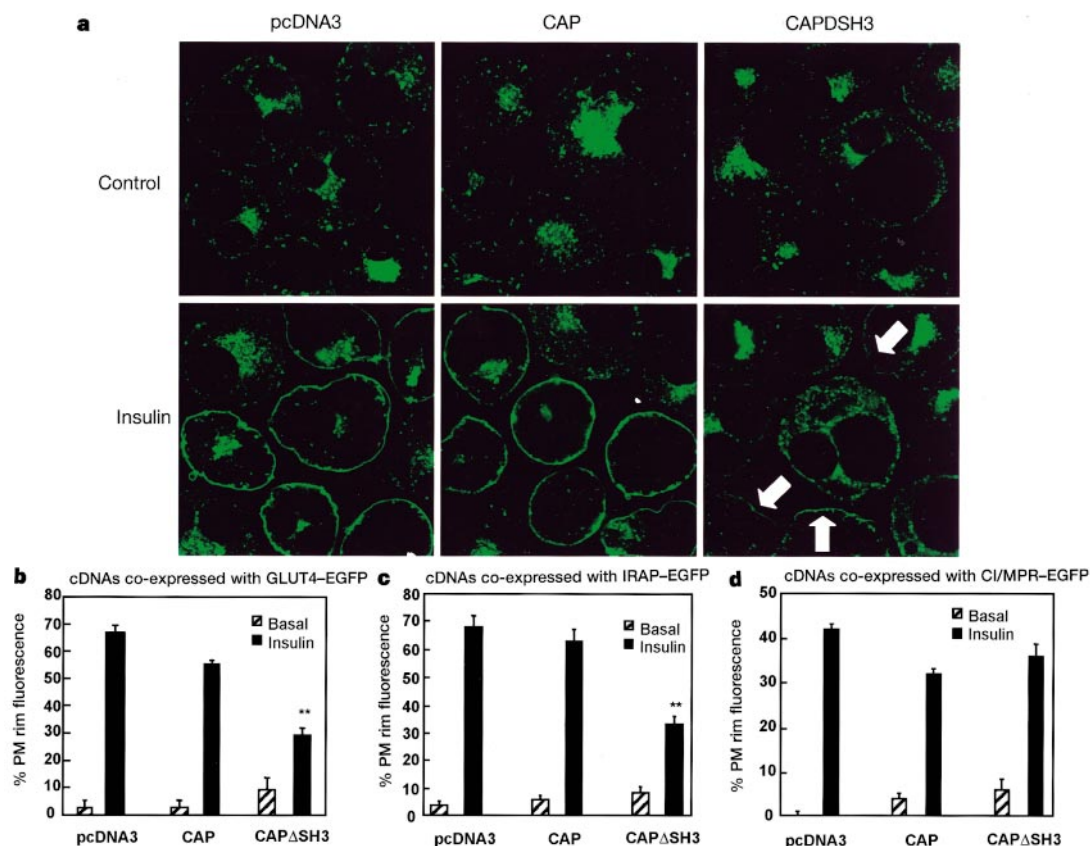


Figure 5 CAP Δ SH3 blocks insulin-stimulated GLUT4 and IRAP translocation to the plasma membrane, without effecting translocation of CI/MPR. **a**, Differentiated 3T3-L1 adipocytes were transfected with GLUT4–EGFP plus vector, Flag–CAP or Flag–CAP Δ SH3, and allowed to recover for 18 h. The cells were treated without or with 100 nM insulin for 30 min. Cells were fixed and fluorescence visualized by confocal microscopy. This is a representative collection of 6–9 cell images per field. **b**, Number of

GLUT4–EGFP transfected cells displaying visually detectable plasma membrane (PM) rim fluorescence. **c**, **d**, Number of EGFP–IRAP- and CI/MPR–EGFP-transfected cells displaying visually detectable rim fluorescence. These data were obtained by blind counting of more than 75 cells from three independent experiments. Double asterisk: significant difference, $P < 0.01$, $n = 3$.

The stimulation of glucose uptake and utilization by insulin is dependent upon the translocation of the insulin-responsive glucose transporter GLUT4 from intracellular storage sites to the plasma membrane. To investigate how CAPΔSH3 blocks glucose transport, the translocation of GLUT4 was directly examined. We electroporated 3T3-L1 adipocytes with pcDNA3, CAP or CAPΔSH3 together with a construct encoding an enhanced green fluorescent protein fusion of GLUT4 (GLUT4-EGFP; Fig. 5a). As expected, insulin stimulated the translocation of GLUT4-EGFP to the plasma membrane in cells transfected with pcDNA3 or CAP, as detected by the presence of a continuous rim of cell-surface EGFP fluorescence. In contrast, expression of CAPΔSH3 markedly inhibited GLUT4-EGFP translocation to the cell surface in response to insulin. Expression of CAPΔSH3 not only decreased the number of cells displaying GLUT4-EGFP at the cell surface, but also decreased the plasma membrane intensity in those cells that still had observable rim fluorescence (arrows). Quantification of the number of transfected cells displaying any visual rim fluorescence indicated that following pcDNA3 or wild-type CAP transfection, about 65% of the cells displayed insulin-stimulated translocation, which was reduced to less than 30% in cells transfected with CAPΔSH3.

The insulin-responsive aminopeptidase (IRAP) is localized in the same compartments as GLUT4 in adipocytes, and undergoes an identical pattern of insulin-stimulated translocation^{13,14}. As expected, expression of CAPΔSH3 blocked insulin-stimulated translocation of EGFP-IRAP, in a manner identical to that observed with EGFP-GLUT4 (Fig. 5c). Insulin also stimulates the net exocytosis of the cation-independent mannose-6-phosphate/IGF-2 receptor (CI/MPR). This protein is located in compartments distinct from the GLUT4/IRAP-containing vesicles, and travels between the trans-Golgi network and the late endosomes in the absence of insulin¹⁵⁻¹⁷. Expression of CAPΔSH3 had no significant effect on the insulin-stimulated translocation of a CI/MPR-EGFP fusion protein (Fig. 5d), in contrast to its effects on GLUT4 and IRAP.

As the determination of translocation in these intact cell assays relies on the qualitative strength of plasma membrane rim fluorescence from overexpressed EGFP-fusion proteins, we also examined the translocation of the endogenous GLUT4 and GLUT1 proteins in isolated plasma membrane sheets from microinjected 3T3-L1 adipocytes (Fig. 6). To identify the plasma membrane sheets derived from the microinjected cells, the wild-type CAP and CAPΔSH3 complementary DNAs were microinjected into the cells with the maltose-binding protein (MBP)-Ras cDNA¹⁸. In the absence of insulin, microinjection of MBP-Ras and CAPΔSH3 resulted in the identification of four MBP-Ras positive sheets in the fields shown in Fig. 6 (arrows).

In the absence of insulin, GLUT4 remained intracellular, with low levels of plasma membrane GLUT4 in both the microinjected and surrounding non-microinjected cells. Insulin stimulation resulted in the appearance of GLUT4 at the plasma membrane in the non-microinjected cells, but not in the cells microinjected with CAPΔSH3. In contrast, microinjection of cells with the wild-type CAP cDNA had no significant effect on GLUT4 translocation, compared with the surrounding non-microinjected cells. Quantification of these data showed that insulin stimulated the translocation of GLUT4 in $87 \pm 6\%$ of the cells microinjected with wild-type CAP. In contrast, expression of CAPΔSH3 resulted in a marked reduction, with only $28 \pm 10\%$ of the plasma membrane sheets displaying any detectable GLUT4 translocation.

Like CI/MPR, the GLUT1 glucose transporter resides in vesicle compartments distinct from GLUT4, and undergoes modest insulin-stimulated plasma membrane translocation¹⁹. To assess the specificity of CAPΔSH3, we examined the translocation of the endogenous GLUT1 protein in microinjected 3T3-L1 adipocytes (see Supplementary Information Fig. 4). Insulin modestly stimulated the translocation of GLUT1. However, this effect of insulin was not impaired by the microinjection of either wild-type CAP or

CAPΔSH3 cDNAs. Under these conditions, $94 \pm 2\%$ and $87 \pm 6\%$ of the cells microinjected with wild-type CAP and CAPΔSH3, respectively, displayed insulin-stimulated GLUT1 translocation. Together, these data show that the CAPΔSH3 mutant has a dominant-interfering effect on insulin-stimulated translocation of GLUT4/IRAP-containing vesicles to the plasma membrane, without affecting general exocytosis or other insulin-regulated vesicle trafficking processes.

The stimulation of glucose transport by insulin is a complex, multistep process involving the movement of intracellularly sequestered GLUT4-containing vesicles to the cell surface, followed by docking and fusion with the plasma membrane¹⁹. The activation of PI(3)K and its downstream kinases AKT and/or protein kinase Cζ/λ (PKCζ/λ) are required for this process, as inhibitors and dominant-negative mutants of PI(3)K, AKT and PKCζ/λ can completely block insulin-stimulated glucose transport²⁰. There is compelling evidence, however, that activation of PI(3)K-dependent pathways is not sufficient to induce glucose transport. Other growth factors and adhesion molecules that can activate PI(3)K and its downstream kinases have no effect on glucose transport or GLUT4 translocation²¹⁻²³. Furthermore, cell-permeable derivatives of phosphatidylinositol 3,4,5-trisphosphate can stimulate GLUT4 translocation only when cells are pre-treated with insulin²⁴. These data suggest that insulin must generate at least two independent signals to stimulate glucose transport, one dependent on and another independent of PI(3)K.

What is the second signal needed for insulin-stimulated glucose transport? Our data indicate that this pathway is probably initiated by the phosphorylation of Cbl and its subsequent segregation into lipid raft subdomains of the plasma membrane. This process is independent of wortmannin (data not shown) and requires the adapter protein CAP². CAP expression and Cbl phosphorylation correlate well with insulin sensitivity²⁵. The appearance of phosphorylated Cbl in the triton-insoluble fraction, its association with the src kinase fyn, and the subsequent phosphorylation of the caveolar protein caveolin are specifically stimulated by insulin. Insulin-stimulated Cbl tyrosine phosphorylation is observed only in insulin-sensitive cells and tissues in which CAP is also expressed^{1,3,11}. Although inhibition of Cbl phosphorylation in the triton-insoluble complexes with the CAPΔSH3 mutant does not interfere with PI(3)K-dependent and ras-dependent signalling, disruption of this association markedly attenuates insulin-stimulated GLUT4 translocation and glucose uptake. Whether the GLUT4 protein itself resides in the triton-insoluble microdomains of the plasma membrane remains uncertain²⁶⁻²⁸. However, it is possible that the association of the CAP-Cbl complex with other proteins in lipid rafts may lead to the segregated generation of a specific signal, perhaps directing fusion of the GLUT4 vesicle to the plasma membrane in a domain near the lipid raft. □

Methods

Cell culture and transfection of HEK293T cells and 3T3-L1 adipocytes

HEK293T cells were maintained in Dulbecco's Eagle Medium (DMEM) containing 10% fetal bovine serum at 37 °C and 5% CO₂. HEK293T cells were transfected with a mammalian CaPO₄ transfection kit as described by the manufacturer (Stratagene). 3T3-L1 preadipocytes were cultured in DMEM containing 25 mM glucose, 10% calf serum at 37 °C and 5% CO₂. Confluent cultures were induced to differentiate into adipocytes as described²⁹. 3T3-L1 adipocytes were transiently transfected by electroporation as described². The cells were then allowed to adhere to collagen-coated tissue culture dishes for 30-48 h before serum starvation and insulin treatment. The average efficiency of transfection was 50% as evaluated by β-galactosidase staining.

In vitro protein associations

The Flag-CAP expression construct was made as described². We constructed Flag-epitope-tagged CAPΔSH3 by restriction digestion, removal of the SH3 domains of the CAP cDNA and religation of the CAP expression construct. CAPΔSH3 does not associate with Cbl or the insulin receptor. We lysed HEK293T cells overexpressing CAP or CAPΔSH3 in HNTG buffer containing protease inhibitors and sodium orthovanadate²⁹. Lysates from HEK293T cells were incubated with GST alone or GST-flotillin immobilized

on glutathione agarose beads for 1 h at 4 °C. Beads were then washed extensively with HNTG and retained proteins were eluted in 4× Laemmli sample buffer, heated for 5 min at 100 °C, separated by SDS–PAGE and immunoblotted with either a Flag antibody (Stratagene) or a polyclonal c-Cbl antibody (Santa Cruz).

GLUT4-EGFP, IRAP-EGFP and CI/M6PR-EGFP translocation assays

We constructed all EGFP fusion proteins and performed translocation assays as described⁹. Briefly, 3T3-L1 adipocytes were electroporated with 50 µg of the EGFP-tagged plasmid and 200 µg of empty vector, CAP or CAPASH3 constructs. Adipocytes were replated, allowed to recover overnight and stimulated with or without 100 nM insulin for 30 min. The cells were then fixed and EGFP fluorescence was analysed by confocal immunofluorescence microscopy.

Single cell microinjection and plasma membrane sheet isolation

We grew 3T3-L1 adipocytes on 35-mm tissue culture dishes. Before microinjection, the medium was changed to Lebovitz's L-15 medium containing 0.1% bovine serum albumin and microinjection was carried out as described³⁰. The cells were allowed to recover for 24 h and plasma membrane sheets were prepared as described¹⁸. The bound plasma membranes sheets were then subjected to confocal fluorescent microscopy.

Metabolic assays

We determined 2-deoxyglucose uptake, glycogen synthesis and lipogenesis as described²⁹.

Immunoprecipitations and immunoblotting

After insulin treatment, we washed electroporated adipocytes with PBS and lysed them in MBS buffer containing protease inhibitors and sodium orthovanadate. Lysed cells were kept at 4 °C to prevent degradation of caveolar complexes. The triton-insoluble pellet fraction was generated as described¹¹. This triton-insoluble pellet fraction was solubilized in SOL buffer containing protease inhibitors and sodium orthovanadate as described¹¹. Lysates and immunoprecipitations were resolved by SDS–PAGE and transferred to nitrocellulose for immunoblot analysis.

Immunoprecipitations were performed using a polyclonal c-Cbl antisera (Santa Cruz). We used monoclonal 4G10 anti-phosphotyrosine antibodies (Upstate Biotechnology) to blot for IR, IRS-1 and Cbl tyrosine phosphorylation. Polyclonal phospho-AKT and phospho-MAPK antibodies (New England Biolabs) were used to probe for activated AKT and MAP kinase. Monoclonal caveolin I antibodies (Transduction Labs) were used to evaluate protein content of the resolved TIP fractions. We used horseradish peroxidase (HRP)-conjugated secondary antibodies and ECL chemiluminescence for immunoblot detection.

Received 13 March; accepted 16 May 2000.

- Ribon, V. & Saltiel, A. R. Insulin stimulates tyrosine phosphorylation of the proto-oncogene product of c-Cbl in 3T3-L1 adipocytes. *Biochem. J.* **324**, 839–845 (1997).
- Ribon, V., Printen, J. A., Hoffman, N. G., Kay, B. K. & Saltiel, A. R. A novel, multifunctional c-Cbl binding protein in insulin receptor signaling in 3T3-L1 adipocytes. *Mol. Cell. Biol.* **18**, 872–879 (1998).
- Mastick, C. C., Brady, M. J. & Saltiel, A. R. Insulin stimulates the tyrosine phosphorylation of caveolin. *J. Cell Biol.* **129**, 1523–1531 (1995).
- Printen, J. A., Brady, M. J. & Saltiel, A. R. PTG, a protein phosphatase 1-binding protein with a role in glycogen metabolism. *Science* **275**, 1475–1478 (1997).
- Bickel, P. E. *et al.* Flotillin and epidermal surface antigen define a new family of caveolae-associated integral membrane proteins. *J. Biol. Chem.* **272**, 13793–13802 (1997).
- Anderson, R. G. The caveolae membrane system. *Annu. Rev. Biochem.* **67**, 199–225 (1998).
- Harder, T. & Simons, K. Caveolae, DIGs, and the dynamics of sphingolipid-cholesterol microdomains. *Curr. Opin. Cell Biol.* **9**, 534–542 (1997).
- Volonte, D. *et al.* Flotillins/cavatellins are differentially expressed in cells and tissues and form a hetero-oligomeric complex with caveolins in vivo. Characterization and epitope-mapping of a novel flotillin-1 monoclonal antibody probe. *J. Biol. Chem.* **274**, 12702–12709 (1999).
- Thurmond, D. C. *et al.* Regulation of insulin-stimulated GLUT4 translocation by Munc18c in 3T3L1 adipocytes. *J. Biol. Chem.* **273**, 33876–33883 (1998).
- Gustavsson, J. *et al.* Localization of the insulin receptor in caveolae of adipocyte plasma membrane. *FASEB J.* **13**, 1961–1971 (1999).
- Mastick, C. C. & Saltiel, A. R. Insulin-stimulated tyrosine phosphorylation of caveolin is specific for the differentiated adipocyte phenotype in 3T3-L1 cells. *J. Biol. Chem.* **272**, 20706–20714 (1997).
- Corvera, S. & Czech, M. P. Direct targets of phosphoinositide 3-kinase products in membrane traffic and signal transduction. *Trends Cell Biol.* **8**, 442–446 (1998).
- Keller, S. R., Scott, H. M., Mastick, C. C., Aebersold, R. & Lienhard, G. E. Cloning and characterization of a novel insulin-regulated membrane aminopeptidase from Glut4 vesicles. *J. Biol. Chem.* **270**, 23612–23618 (1995); erratum, *ibid* **270**, 30236 (1995).
- Kandror, K. V. & Pilch, P. F. gp160, a tissue-specific marker for insulin-activated glucose transport. *Proc. Natl Acad. Sci. USA* **91**, 8017–8021 (1994).
- Oka, Y., Mottola, C., Oppenheimer, C. L. & Czech, M. P. Insulin activates the appearance of insulin-like growth factor II receptors on the adipocyte cell surface. *Proc. Natl Acad. Sci. USA* **81**, 4028–4032 (1984).
- Wardzala, L. J., Simpson, I. A., Rechler, M. M. & Cushman, S. W. Potential mechanism of the stimulatory action of insulin on insulin-like growth factor II binding to the isolated rat adipose cell. Apparent redistribution of receptors cycling between a large intracellular pool and the plasma membrane. *J. Biol. Chem.* **259**, 8378–8383 (1984).
- Malide, D. & Cushman, S. W. Morphological effects of wortmannin on the endosomal system and GLUT4-containing compartments in rat adipose cells. *J. Cell Sci.* **110**, 2795–2806 (1997).

- Min, J. *et al.* Synip: a novel insulin-regulated syntaxin 4-binding protein mediating GLUT4 translocation in adipocytes. *Mol. Cell* **3**, 751–760 (1999).
- Pessin, J. E., Thurmond, D. C., Elmendorf, J. S., Coker, K. J. & Okada, S. Molecular basis of insulin-stimulated GLUT4 vesicle trafficking. Location! Location! Location! *J. Biol. Chem.* **274**, 2593–2596 (1999).
- Czech, M. P. & Corvera, S. Signaling mechanisms that regulate glucose transport. *J. Biol. Chem.* **274**, 1865–1868 (1999).
- Wiese, R. J., Mastick, C. C., Lazar, D. F. & Saltiel, A. R. Activation of mitogen-activated protein kinase and phosphatidylinositol 3'-kinase is not sufficient for the hormonal stimulation of glucose uptake, lipogenesis, or glycogen synthesis in 3T3-L1 adipocytes. *J. Biol. Chem.* **270**, 3442–3446 (1995).
- Isakoff, S. J. *et al.* The inability of phosphatidylinositol 3-kinase activation to stimulate GLUT4 translocation indicates additional signaling pathways are required for insulin-stimulated glucose uptake. *Proc. Natl Acad. Sci. USA* **92**, 10247–10251 (1995).
- Guilherme, A. & Czech, M. P. Stimulation of IRS-1-associated phosphatidylinositol 3-kinase and Akt/protein kinase B but not glucose transport by beta1-integrin signaling in rat adipocytes. *J. Biol. Chem.* **273**, 33119–33122 (1998).
- Jiang, T. *et al.* Membrane-permeant esters of phosphatidylinositol 3,4,5-trisphosphate. *J. Biol. Chem.* **273**, 11017–11024 (1998).
- Ribon, V., Johnson, J. H., Camp, H. S. & Saltiel, A. R. Thiazolidinediones and insulin resistance: peroxisome proliferator-activated receptor gamma activation stimulates expression of the CAP gene. *Proc. Natl Acad. Sci. USA* **95**, 14751–14756 (1998).
- Kandror, K. V., Stephens, J. M. & Pilch, P. F. Expression and compartmentalization of caveolin in adipose cells: coordinate regulation with and structural segregation from GLUT4. *J. Cell Biol.* **129**, 999–1006 (1995).
- Scherer, P. E. *et al.* Induction of caveolin during adipogenesis and association of GLUT4 with caveolin-rich vesicles. *J. Cell Biol.* **127**, 1233–1243 (1994).
- Gustavsson, J., Parpal, S. & Stralfors, P. Insulin-stimulated glucose uptake involves the transition of glucose transporters to a caveolae-rich fraction within the plasma membrane: implications for type II diabetes. *Mol. Med.* **2**, 367–372 (1996).
- Lazar, D. F. *et al.* Mitogen-activated protein kinase kinase inhibition does not block the stimulation of glucose utilization by insulin. *J. Biol. Chem.* **270**, 20801–20807 (1995).
- Elmendorf, J. S., Chen, D. & Pessin, J. E. Guanosine 5'-O-(3-thiotriphosphate) (GTPγS) stimulation of GLUT4 translocation is tyrosine kinase-dependent. *J. Biol. Chem.* **273**, 13289–13296 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

J.E.P. is supported by the NIH. P.E.B. is supported by the American Diabetes Association and awards from the Washington University Clinical Nutrition Research Unit and Diabetes Research Training Center.

Correspondence and requests for materials should be addressed to A.R.S. (e-mail: Alan.Saltiel@wl.com).

Activation of heat-shock response by an adenovirus is essential for virus replication

Jolanta B. Glotzer*, Mediha Saltik*, Susanna Chiocca†, Anne-Isabelle Michou*, Pope Moseley‡ & Matt Cotten*

* Institute for Molecular Pathology, Dr Bohr-Gasse 7, A-1030 Vienna, Austria
 † European Institute of Oncology, Department of Experimental Oncology, 20141 Milan, Italy
 ‡ The University of New Mexico, Department of Medicine, Albuquerque, New Mexico 87131-5271, USA

Successful viral infection requires viruses to redirect host biochemistry to replicate the viral genome, and produce and assemble progeny virions. Cellular heat-shock responses, which are characterized as elevation and relocation of heat-shock proteins, occur during replication of many viruses^{1–7}. Such responses might be host reactions to the synthesis of foreign protein, or might be irrelevant consequences of the viral need to activate transcription. Alternatively, as heat-shock proteins can facilitate protein folding^{8,9}, activating a heat-shock response might be a specific virus function ensuring proper synthesis of viral proteins and virions. It is not possible to determine whether heat-shock response is essential for virus replication, because the implicated

viral genes (such as Ad5 E1A, ref. 10) also control other essential replication steps. Here we report that expression of Gam1, a protein encoded by the avian virus CELO (ref. 11), elevates and relocalizes hsp70 and hsp40. Gam1-negative CELO is replication-defective; however, Gam1 function can be partially replaced by either heat shock or forced hsp40 expression. Thus, an essential function of Gam1 during virus replication is to activate host heat-shock responses with hsp40 as a primary target.

Gam1, encoded by the avian adenovirus CELO, was identified in an anti-apoptosis screen¹¹. The nuclear location of Gam1 suggested that the protein might influence the expression of genes whose products could modulate apoptosis. Among these, hsp70 expression has been correlated with increased cell survival under stress^{12,13}, and thus it was logical to examine the influence of Gam1 on hsp70 expression.

We constructed an E1-defective adenovirus 5 vector (AdGam1) to direct substantial levels of Gam1 expression (Fig. 1a, b). To determine whether Gam1 expression influences hsp70 protein levels, we infected A549 cells with either AdGam1 or control adenoviruses encoding nuclear-targeted β -galactosidase (Ad β gal), luciferase (AdLuc) or enhanced green fluorescent protein (AdEGFP). The virus-encoded transgene products were expressed to various levels and detectable by Coomassie staining (Fig. 1a, b, left panels). A substantial increase in hsp70 protein occurred in cells that expressed Gam1 (Fig. 1a, b, right panel) or that were exposed to heat shock (Fig. 1a, right panel), but not in cells that were infected with the same amount of control adenoviruses Ad β gal, AdLuc or AdEGFP in the absence of heat shock (Fig. 1a, b). This result shows that vector infection itself, background gene expression per se do not alter hsp70 levels. Gam1-directed hsp70 induction was similar to that obtained

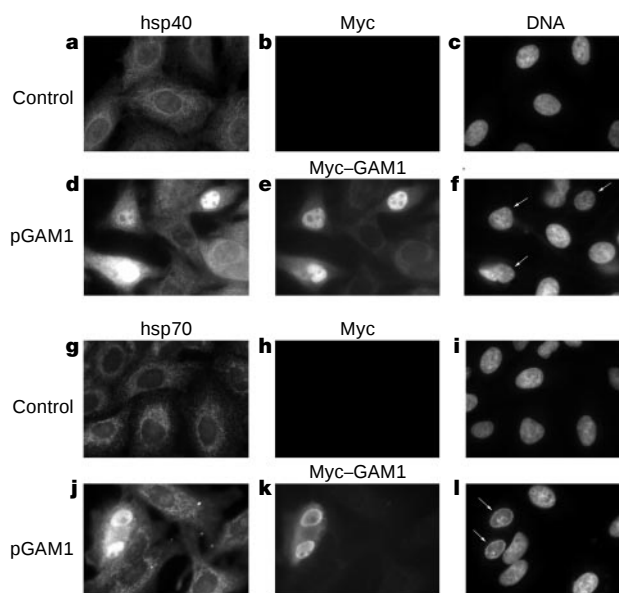


Figure 2 hsp40 and hsp70 are upregulated by Gam1 expressed from a transfected plasmid. Immunofluorescence analysis of A549 cells non-transfected (control, **a–c, g–i**) or transfected with pGAM1 (CMV immediate early promoter driving a Myc–Gam1 cDNA, **d–f, j–l**) stained for hsp40 (**a, d**) or hsp70 (**g, j**) and Myc for Myc-tagged Gam1 (**b, e, h, k**). Nuclei are counterstained with Hoechst dye (**c, f, i, l**). In **f** and **l**, arrows indicate cells with strong nuclear Gam1 expression and upregulation, and nuclear accumulation of hsp40 and hsp70, respectively.

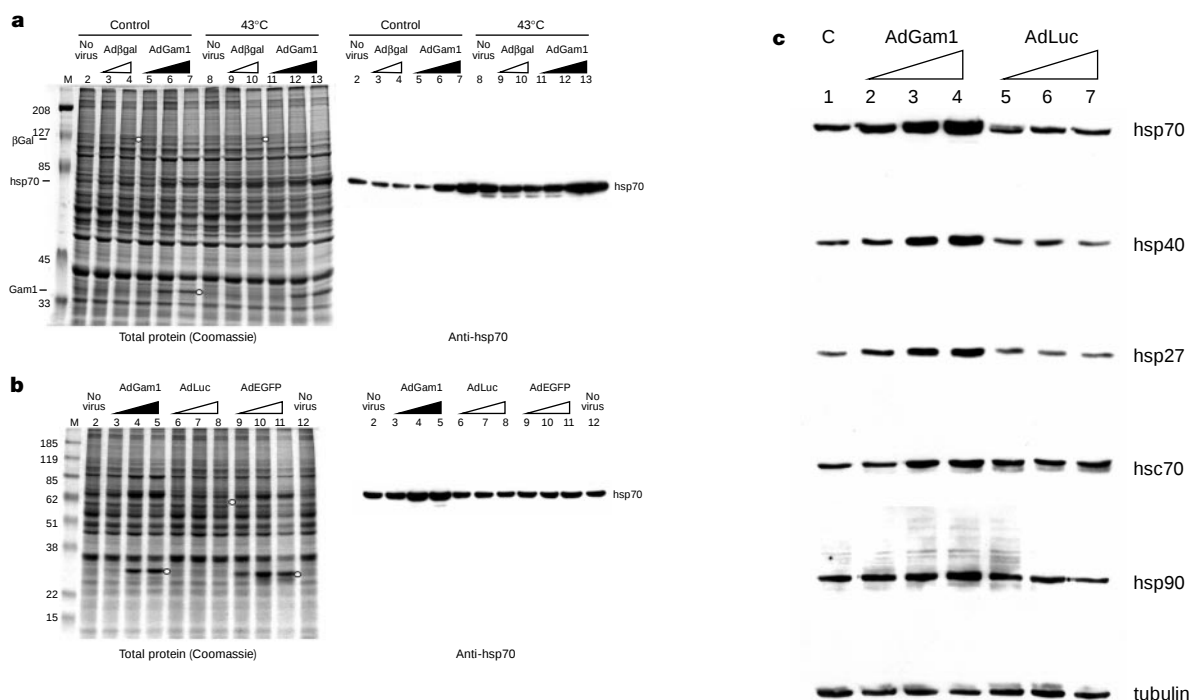


Figure 1 Gam1 expression elevates heat-shock protein levels. **a**, A549 cells were infected with the indicated adenovirus; 24 h later, heat shock (43 °C, 90 min) was applied as indicated; 48 h after infection, extracts were resolved by SDS–PAGE and stained for total protein (Coomassie blue) or immunoblotted for hsp70. Circles indicate viral transgene product. Lanes, 2–7, no heat shock; lanes, 8–13, heat shock. Non-infected cells (lanes 2, 8); infection with Ad β gal at 1,000 (lanes 3, 9) or 10,000 particles per cell (lanes 4, 10); with AdGam1 at 100 (lanes 5, 11), 1,000 (lanes 6, 12) or 10,000 particles

per cell (lanes 7, 13). **b**, A549 cells infected with the indicated adenovirus at 1,000, 3,000 or 10,000 particles per cell; with extracts prepared 48 h after infection and analysed as above. Lanes 2, 12; non-infected cells; lanes 3–5, infection with AdGam1; lanes 6–8, with AdLuc; lanes 9–11, with AdEGFP. **c**, A549 cells infected with AdGam1 or AdLuc at 1,000, 3,000 or 10,000 particles per cell with extracts prepared 48 h post-infection and immunoblotted for hsp70, hsp40, hsp27, hsc70, hsp90 or tubulin.

with heat shock (Fig. 1b, right panel). AdGam1 infection but not AdLuc infection also led to increases of hsp40 and hsp27 but not to significant increases in hsc70 or hsp90 α (Fig. 1c) hsp70 and hsp40 move to the nucleus following heat shock^{14,15}. Non-transfected, non-infected A549 cells present predominantly cytoplasmic hsp40 and hsp70 (Fig. 2a–c, g–i). Gam1 expression from a transfected plasmid causes an elevation and nuclear relocalization of both hsp40 (Fig. 2d–f) and hsp70 (Fig. 2j–l). Infection with AdGam1 also led to elevation and nuclear relocation of hsp40 (Fig. 3a–c) and hsp70 (Fig. 3g–i). Neither transfection with a control plasmid (pEGFP; data not shown) nor infection of cells with a control AdEGFP expressing comparable levels of protein (Fig. 1b) altered hsp40 or hsp70 levels or localization (Fig. 3d–f, j–l). hsp40 and hsp70 levels also increased during CELO replication (Fig. 4).

To examine the requirement for Gam1 in CELO replication, we prepared a Gam1-negative CELO genome bearing a luciferase gene (CELOdG; see Methods). The CELOdG genome yielded replicating virus when complemented with AdGam1 (Fig. 5a), but not with a control AdEGFP (Fig. 5d). Thus, the Gam1 gene is required for CELO growth in LMH cells. In chicken embryos, wild-type CELO (CELOwt) produces a maximum virus yield with inocula of 40 particles and higher; full replication of CELOdG is only observed with inocula of 4×10^7 particles and above (data not shown), demonstrating the requirement for Gam1 during *in vivo* virus replication.

To determine whether activating a heat-shock response is a necessary function of Gam1, we tested heat shock for its ability to replace Gam1 in CELO replication. When the genome of CELOdG was transfected into LMH cells and the culture exposed to 45 °C for 90 min, viral growth was observed (data not shown). We used heat-shock complementation to amplify CELOdG, and then compared purified CELOdG virus with CELOwt in a single cycle of virus replication by measuring the production of capsid proteins (Fig. 5b) or the production of virus particles capable of delivering a luciferase gene (Fig. 5c). The production of CELOdG, which is clearly

defective in non-heated cultures (Fig. 5b, lane 3; 5c, lane 2), was stimulated by heat shock (Fig. 5b, lanes 6, 9 and 12; and 5c, lanes 4, 6 and 8), with the greatest stimulation observed when the heat shock was applied 2 h before virus infection (Fig. 5b, lane 12; 5c, lane 8).

The heat shock used for complementation might influence a variety of cellular proteins in addition to hsp40 and hsp70. To determine whether either or both hsp40 and hsp70 are essential Gam1 targets for CELO replication, we prepared recombinant adenoviruses directing the expression of these two heat-shock proteins (Adhsp40, Adhsp70) and tested them for complementing CELOdG growth in LMH cells. Overexpression of hsp40 but not hsp70 supports CELOdG growth (Fig. 5d). CELOdG genomes were also constructed that contained either hsp40 or hsp70 expression cassettes in place of the Gam1 gene (CELOdGhsp40, CELOdGhsp70). Introduction of the CELOdGhsp40 genome into LMH cells yielded replicating virus; the CELOdGhsp70 genome, like the parental CELOdG, was not viable (two independent genomes were tested; data not shown). In a single cycle of replication, CELOdGhsp40 was found to replicate at roughly one-twentieth of the level of CELOwt, whereas CELOdG growth was not detectable (Fig. 5e). Thus, hsp40 overexpression alone is sufficient partially to replace Gam1 function.

We have shown that the CELO Gam1 protein increases the cellular levels of hsp70 and hsp40 and relocates these proteins to the nucleus. Gam1 is required for virus replication, and, moreover, the requirement for Gam1 in viral replication can be partially met either by heat shock or by overexpression of hsp40. Our data support the idea that the induction of a heat-shock response is an essential viral function for viral replication and is not merely a cellular adaptive response to infection. The relocalization of hsp40 and hsp70 to the nucleus by Gam1 probably facilitates the protein reorganization required early in virus replication. Heat-shock induction is observed during the replication of several viruses including herpes simplex, cytomegalovirus and adenovirus, suggesting that heat-shock protein function in viral replication may serve a common mechanism for all of these viruses. DnaK

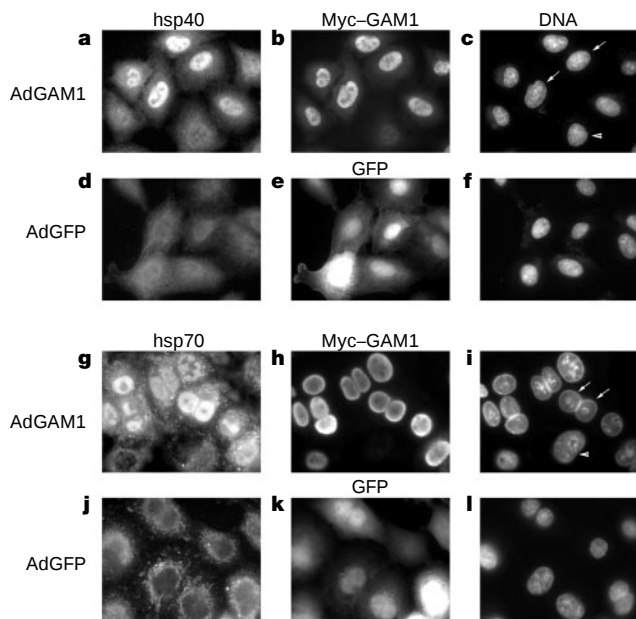


Figure 3 Gam1 relocates hsp40 and hsp70 to the nucleus. Immunofluorescence analysis of A549 cells infected with AdGam1 (a–c, g–i) or AdGFP (d–f, j–l) at 10,000 particles per cell. Staining with anti-hsp40 (a, d) anti-Myc for Myc-tagged Gam1 (b, h), or with anti-hsp70 (g, j) and GFP expression (e, k); nuclei were counterstained with Hoechst dye (c, f, i, l). In c and i, arrows indicate cells with strong Gam1 expression and nuclear accumulation of hsp40 and hsp70; arrowheads indicate cells with weak (c) or no detectable Gam1 expression (i) and no nuclear accumulation of hsp40 or hsp70.

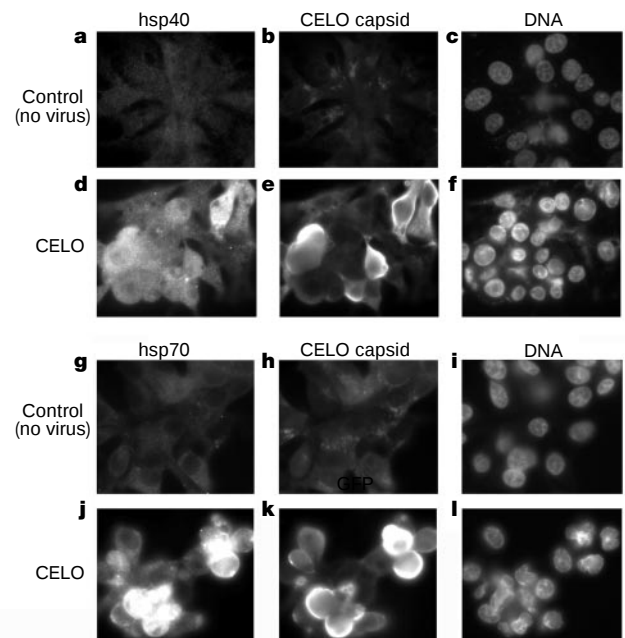


Figure 4 hsp40 and hsp70 are upregulated during CELO infection. Immunofluorescence analysis of LMH cells that were non-infected (a–c; g–i) or infected with wild-type CELO (100 particles per cell; d–f; j–l). Cells were labelled with anti-hsp40 (a, d) or anti-hsp 70 (g, j), and with anti-CELO capsid (b, e, h, k) antibodies. Nuclei were counterstained with Hoechst dye (c, f, i, l).

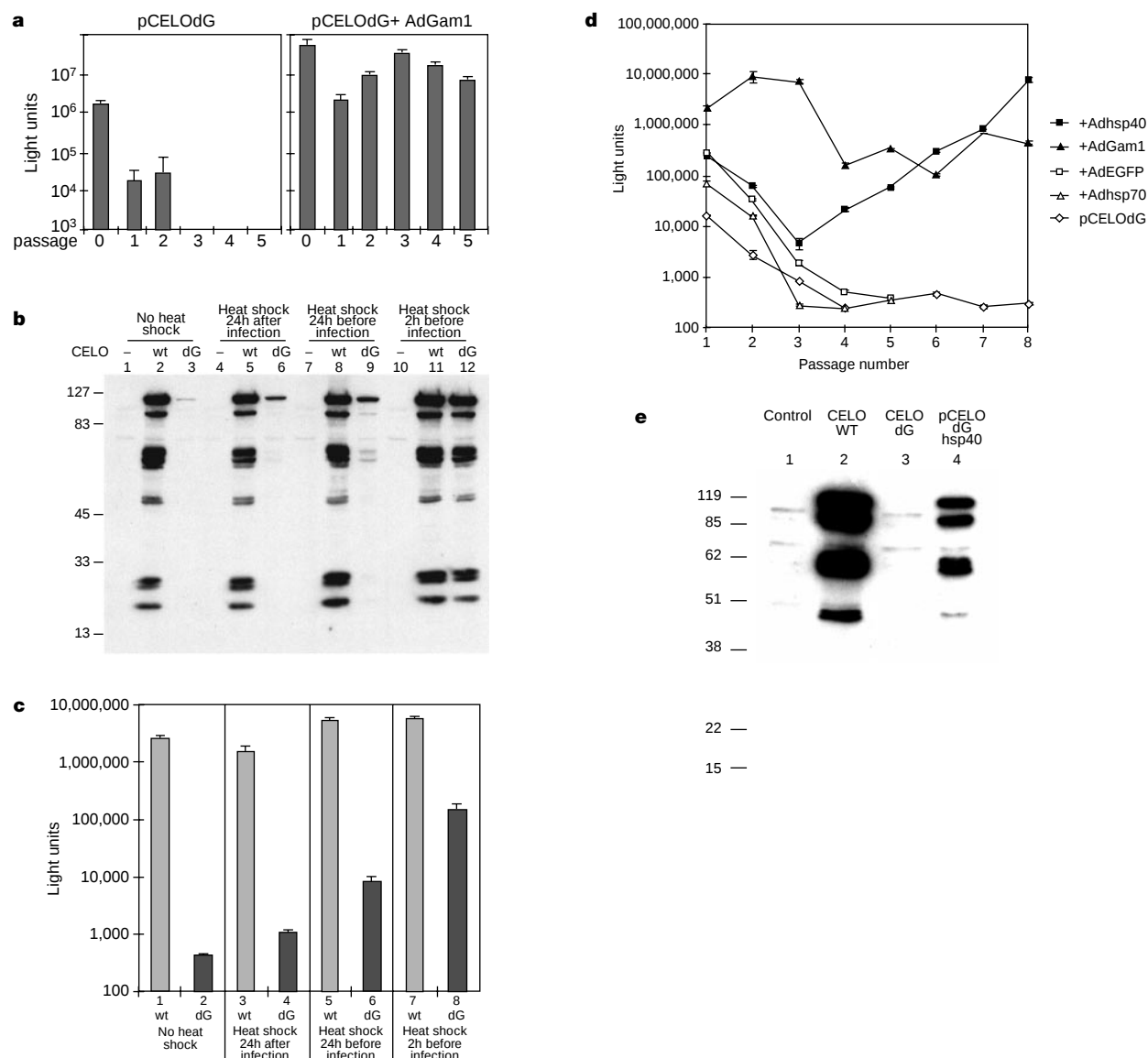


Figure 5 Gam1 is required for CELO replication and can be replaced by heat shock or hsp40. **a**, Left, replication of CELOdG alone; right, replication of CELOdG plus AdGam1. Luciferase activity (mean $\pm$ s.d. of three cultures) is indicated for each passage. **b**, **c**, LMH cells were infected with 5 particles per cell of CELOwt or CELOdG. Cells were incubated without heat shock (**b**, lanes 1–3; **c**, lanes 1, 2); with a 90-min, 45 °C heat-shock 24 h after infection (**b**, lanes 4–6; **c**, lanes 3, 4); 24 h before infection (**b**, lanes 7–9; **c**, lanes 5, 6); or 2 h before infection (**b**, lanes 10–12; **c**, lanes 7, 8). After 4 days at 37 °C, lysates were analysed by immunoblotting for CELO capsid proteins (**b**) or for virus capable of transducing

the luciferase gene into LMH cells (**c**). **d**, LMH cells were infected with 1 virus particle per cell of CELOdG either alone or with 10,000 particles per cell of AdEGFP, AdGam1, Adhsp40 or Adhsp70. For each passage, lysates were prepared 5 days after infection, luciferase was measured (mean $\pm$ s.d. of three results), and a fresh monolayer of LMH cells was infected (plus the indicated second virus). **e**, LMH cells were infected with 5 particles per cell of CELOwt, CELOdG or CELOdGhsp40. After 5 days at 37 °C, equal quantities of lysate protein were analysed by immunoblotting for virus capsid proteins. Lane 1, non-infected cells; lane 2, CELOwt; lane 3, CELOdG; lane 4, CELOdGhsp40.

and DnaJ, the bacterial hsp70 and hsp40 homologues, were originally identified as host factors essential for bacteriophage lambda replication (reviewed in ref. 16). hsp40 and hsp70 stimulate papillomavirus replication¹⁷ and DNA tumour virus T antigens have been shown to possess DnaJ-like activity that is important for remodeling nuclear protein complexes required for viral replication^{18–21}. To date, the CELO virus is unique in having evolved a replication strategy using a single gene product, Gam1, which induces a heat-shock response in the host cells essential for the replication of the virus.

Methods

Modified adenoviruses

For AdGam1, the Gam1 coding sequence (nucleotides 37,391–38,239 in the CELO

genome²²) plus an amino-terminal Myc tag were amplified from pSG9mGam1 (ref. 11) using polymerase chain reaction and transferred into an E1/E3 negative adenovirus 5 genome using homologous recombination in bacteria^{23,24}. The final virus bears an expression unit containing a CMV promoter, the Myc-tagged Gam1 coding sequence and rabbit β -globin intron poly(A) signal embedded in the E1 region. We used similar methods to construct Adhsp40 (containing a human hsp40 cDNA²⁵). An adenovirus bearing a human hsp70 coding sequence²⁶ with a tet-repressible CMV promoter will be described elsewhere. Control E1-negative Ad5 viruses expressing luciferase (AdLuc) and eGFP (AdEGFP)²⁴, or β -galactosidase (AdRSV β gal)²⁷, and the purification of adenovirus on CsCl gradients have been described²⁴.

CELOdG construction

CELO with deleted Gam1 gene (CELOdG) was constructed using a deletion/recombination method²⁴. In brief, a fragment of the CELO genome was manipulated to delete a *Sma*I/*Bgl*II fragment (CELO nucleotides 36,818–37,972) and insert a luciferase expression cassette. This deletion removes the first 602 bp of the 875 bp Gam1 coding sequence. The modified fragment was assembled into a complete CELO genome to generate pCELOdG

(pAIM65). The luciferase expressing, wild-type Gam1 (CELOWt) and the CELO virus purification on CsCl gradients have been described (CELO AIM46; ref. 24). The CELOdGhsp40 and CELOdGhsp70 genomes were constructed by exchanging the CMV/luciferase/ β -globin cassette for CMV/hsp40/ β -globin or CMV/hsp70/ β -globin cassettes.

Analysis of the replication of CELOdG (Fig. 5a) was performed by transfecting the CELOdG genome into LMH cells alone, or followed by infection with 1,000 particles per cell of AdGam1 after 24 h. After 5 days, we collected the cells and assayed them for luciferase activity. We used an additional culture to infect a fresh set of LMH cultures either alone, or with AdGam1. We repeated the cell collection, luciferase assay and passage every 5 days for 5 passages.

Immunoblotting analysis

We lysed A549 cells in lysis buffer (150 mM NaCl, 50 mM Tris, pH 7.5, 5 mM EDTA, 1% NP-40 containing protease inhibitor cocktail; Sigma). Cells were agitated at 4 °C for 30 min, passed through a 25 gauge needle 5 times, sonicated in a bath sonicator for 5 min, and centrifuged at 14,000 r.p.m. (Eppendorf) for 5 min, and the supernatant was used for immunoblotting analysis. Equal quantities of protein (measured by Bradford reagent, Pierce) were resolved by polyacrylamide gel electrophoresis (PAGE), transferred to nitrocellulose and probed with the indicated antiserum preparations (see below). Antibody binding was revealed using the appropriate peroxidase-conjugated secondary antibodies (Dako) and ECL reagents (Amersham).

Immunofluorescence

Cells were plated on glass cover slips (12 × 12 mm) in 6-well dishes 24 h before transfection or infection. Cells were fixed one day after transfection/infection in 4% paraformaldehyde for 15 min, rinsed 3 times with PBS, permeabilized with PBS/0.1% Triton X-100 (PBT) for 15 min, blocked in 5% BSA/PBT for 30–60 min, incubated with primary antibody for 15 min in 5% BSA/PBT, washed 3 times with PBT, incubated with secondary antibody for 15 min in 5% BSA/PBT, washed 2 times with PBT, washed 2 times with PBS, and mounted in 50% glycerol/PBS, 10 mM Tris pH 8.5, 4% *n*-propyl gallate (Sigma); all incubations were done at room temperature. DNA was stained with Hoechst dye. Images were acquired with a cooled CCD camera (Spot II; Diagnostic Instruments) mounted on an Axiovert microscope (Zeiss) equipped with 63×/1.4 lens with filters from Chroma Tech and processed using Adobe Photoshop software.

We used the following antiserum preparations: murine monoclonal 9E10 recognizing the Myc epitope²⁸; murine monoclonal RPN-1197 recognizing hsp70 (Amersham) or goat polyclonal sera recognizing hsp70, hsp40, hsp90 α , hsc70 and hsp27 (Santa Cruz Biotechnology); anti-tubulin (clone DM1A, Sigma); and a rabbit polyclonal directed against total capsid proteins²⁴. We used the following secondary antibodies: DTAF conjugated donkey anti mouse, DTAF conjugated donkey anti rabbit and Cy3 conjugated donkey anti goat (Jackson Laboratories).

We carried out transfections using a double PEI technique as described²⁴. LMH cells²⁹ and A549 cells (ATCC CCL-185) were cultured in DMEM plus 10% FCS (DMEM, 2 mM glutamine, 100 IU penicillin, 100 μ g ml⁻¹ streptomycin and 10% (v/v) fetal calf serum). The 293 cell line³⁰ was cultured in MEMalpha with 10% newborn calf serum.

Received 24 January; accepted 23 June 2000.

- Nevins, J. R. Induction of the synthesis of a 70,000 dalton mammalian heat shock protein by the adenovirus E1A gene product. *Cell* **29**, 913–919 (1982).
- Kao, H. T. & Nevins, J. R. Transcriptional activation and subsequent control of the human heat shock gene during adenovirus infection. *Mol. Cell. Biol.* **3**, 2058–2065 (1983).
- Phillips, B., Abravaya, K. & Morimoto, R. I. Analysis of the specificity and mechanism of transcriptional activation of the human hsp70 gene during infection by DNA viruses. *J. Virol.* **65**, 5680–5692 (1991).
- Santomenna, L. D. & Colberg-Poley, A. M. Induction of cellular hsp70 expression by human cytomegalovirus. *J. Virol.* **64**, 2033–2040 (1990).
- Ohgita, E., Kobayashi, K., Takeshita, K. & Imanishi, J. Biphasic translocation of a 70 kDa heat shock protein in human cytomegalovirus-infected cells. *J. Gen. Virol.* **80**, 63–68 (1999).
- Kobayashi, K. *et al.* Herpes simplex virus-induced expression of 70 kDa heat shock protein (HSP70) requires early protein synthesis but not viral DNA replication. *Microbiol. Immunol.* **38**, 321–325 (1994).
- Collins, P. L. & Hightower, L. E. Newcastle disease virus stimulates the cellular accumulation of stress (heat shock) mRNAs and proteins. *J. Virol.* **44**, 703–707 (1982).
- Mayer, M. P. & Bukau, B. HSP70 chaperone systems: diversity of cellular functions and mechanism of action. *Biol. Chem.* **379**, 261–268 (1998).
- Kelley, W. L. Molecular chaperones: How J domains turn on Hsp70s. *Curr. Biol.* **9**, R305–308 (1999).
- Flint, J. & Shenk, T. Viral transactivating proteins. *Annu. Rev. Genet.* **31**, 177–212 (1997).
- Chiocca, S., Baker, A. & Cotten, M. Identification of a novel anti-apoptotic protein, GAM-1, encoded by the CELO adenovirus. *J. Virol.* **71**, 3168–3177 (1997).
- Gabai, V. L. *et al.* Hsp70 prevents activation of stress kinases. A novel pathway of cellular thermotolerance. *J. Biol. Chem.* **272**, 18033–18037 (1997).
- Mosser, D. D. *et al.* Role of the human heat shock protein hsp70 in protection against stress-induced apoptosis. *Mol. Cell. Biol.* **17**, 5317–5327 (1997).
- Welch, W. J. & Feramisco, J. R. Nuclear and nucleolar localization of the 72,000-dalton heat shock protein in heat-shocked mammalian cells. *J. Biol. Chem.* **259**, 4501–4513 (1984).
- Hattori, H. *et al.* Intracellular localization and partial amino acid sequence of a stress-inducible 40-kDa protein in HeLa cells. *Cell. Struct. Funct.* **17**, 77–86 (1992).
- Polissi, A., Goffin, L. & Georgopoulos, C. The *Escherichia coli* heat shock response and bacteriophage lambda development. *FEMS Microbiol. Rev.* **17**, 159–169 (1995).
- Liu, J. S. *et al.* Human Hsp70 and Hsp40 chaperone proteins facilitate human papillomavirus-11 E1 protein binding to the origin and stimulate cell-free DNA replication. *J. Biol. Chem.* **273**, 30704–30712 (1998).

- Campbell, K. S. *et al.* DnaJ/hsp40 chaperone domain of SV40 large T antigen promotes efficient viral DNA replication. *Genes Dev.* **11**, 1098–1110 (1997).
- Kelley, W. L. & Georgopoulos, C. The T/t common exon of simian virus 40, JC, and BK polyomavirus T antigens can functionally replace the J-domain of the *Escherichia coli* DnaJ molecular chaperone. *Proc. Natl Acad. Sci. USA* **94**, 3679–3684 (1997).
- Stubdal, H. *et al.* Inactivation of pRB-related proteins p130 and p107 mediated by the J domain of simian virus 40 large T antigen. *Mol. Cell. Biol.* **17**, 4979–4990 (1997).
- Srinivasan, A. *et al.* The amino-terminal transforming region of simian virus 40 large T and small t antigens functions as a J domain. *Mol. Cell. Biol.* **17**, 4761–4773 (1997).
- Chiocca, S. *et al.* The complete DNA sequence and genomic organization of the avian adenovirus CELO. *J. Virol.* **70**, 2939–2949 (1996).
- Chartier, C. *et al.* Efficient generation of recombinant adenovirus vectors by homologous recombination in *Escherichia coli*. *J. Virol.* **70**, 4805–4810 (1996).
- Michou, A. -I., Lehrmann, H., Saltik, M. & Cotten, M. Mutational analysis of the avian adenovirus CELO, which provides a basis for gene delivery vectors. *J. Virol.* **73**, 1399–1410 (1999).
- Ohtsuka, K. Cloning of a cDNA for heat-shock protein hsp40, a human homologue of bacterial DnaJ. *Biochem. Biophys. Res. Commun.* **197**, 235–240 (1993).
- Hunt, C. & Morimoto, R. I. Conserved features of eukaryotic hsp70 genes revealed by comparison with the nucleotide sequence of human hsp70. *Proc. Natl Acad. Sci. USA* **82**, 6455–6459 (1985).
- Stratford-Perricaudet, L. D., Makeh, L., Perricaudet, M. & Briand, P. Widespread long-term gene transfer to mouse skeletal muscles and heart. *J. Clin. Invest.* **90**, 626–630 (1992).
- Evan, G. I., Lewis, G. K., Ramsay, G. & Bishop, J. M. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol. Cell. Biol.* **5**, 3610–3616 (1985).
- Kawaguchi, T., Nomura, K., Hirayama, Y. & Kitagawa, T. Establishment and characterization of a chicken hepatocellular carcinoma cell line, LMH. *Cancer Res.* **47**, 4460–4464 (1987).
- Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.* **36**, 59–74 (1977).

Acknowledgements

We thank M. Raff for suggesting the heat-shock complementation and G. Christofori for his comments on the manuscript. We thank K. Ohtsuka for the hsp40 cDNA and S. Fox and R. Morimoto for the hsp70 gene.

Correspondence and requests for materials should be addressed to M.C. (e-mail: cotten@nt.imp.univie.ac.at).

Peroxynitrite reductase activity of bacterial peroxiredoxins

Ruslana Bryk*, Patrick Griffin† & Carl Nathan*

* Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, New York 10021, USA

† Basic Chemistry Analytical Support, Merck Research Laboratories, Rahway, New Jersey 07065, USA

Nitric oxide (NO) is present in soil and air, and is produced by bacteria, animals and plants. Superoxide (O₂⁻) arises in all organisms inhabiting aerobic environments. Thus, many organisms are likely to encounter peroxynitrite (OONO⁻), a product of NO and O₂ that forms at near diffusion-limited rates, and rapidly decomposes upon protonation through isomerization to nitrate (NO₃⁻; ref. 1) while generating hydroxyl radical (·OH) and nitrogen dioxide radical (·NO₂) (refs 2, 3), both more reactive than peroxynitrite's precursors. The oxidative, inflammatory, mutagenic and cytotoxic potential (ref. 4) of peroxynitrite contrasts with the anti-oxidant, anti-inflammatory and tissue-protective properties ascribed to NO itself. Thus, the ability of cells to cope with peroxynitrite is central in determining the biological consequences of NO production. We considered whether cells might be equipped with enzymes to detoxify peroxynitrite. Peroxiredoxins have been identified in most genomes sequenced, but their functions are only partly understood. Here we show that the peroxiredoxin alkylhydroperoxide reductase subunit C (AhpC) from *Salmonella typhimurium* catalytically detoxifies peroxynitrite to nitrite fast enough to forestall the oxidation of bystander molecules such as DNA. Results are similar with peroxiredoxins from *Mycobacterium tuberculosis* and *Helicobacter pylori*. Thus,

peroxynitrite reductase activity may be widespread among bacterial genera.

Enzymatic activities have been demonstrated for only a few of the proteins encoded by several hundred peroxiredoxin sequences, namely, the reduction of alkyl or hydrogen peroxides to alcohols or water⁶. The first peroxiredoxin purified, AhpC, was identified in *Escherichia coli* and *S. typhimurium* along with its flavoprotein partner AhpF as products of a bicistronic operon within the H₂O₂-activatable oxyR regulon⁷. We recently demonstrated that an *ahpCF*-deficient mutant of *S. typhimurium* was markedly sensitized to bactericidal effects of nitrite and *S*-nitrosoglutathione⁸. Resistance was restored by transformation with *ahpC* from *S. typhimurium* or *M. tuberculosis*. Moreover, transfection with *ahpC* from *M. tuberculosis* protected human cells from death caused by nitric oxide synthase-2 (NOS2)⁸. Host defence against experimental tuberculosis depends on production of reactive nitrogen intermediates by NOS2⁹, and antisense-mediated decrease in expression of *ahpC* abolished the virulence of *Mycobacterium bovis*¹⁰. Thus, peroxiredoxins can protect cells from reactive nitrogen intermediates, and we hypothesized that peroxiredoxins do so by acting as peroxynitrite reductases.

We tested this hypothesis using pure *S. typhimurium* AhpC, reagent peroxynitrite, and target molecules susceptible to oxidation.

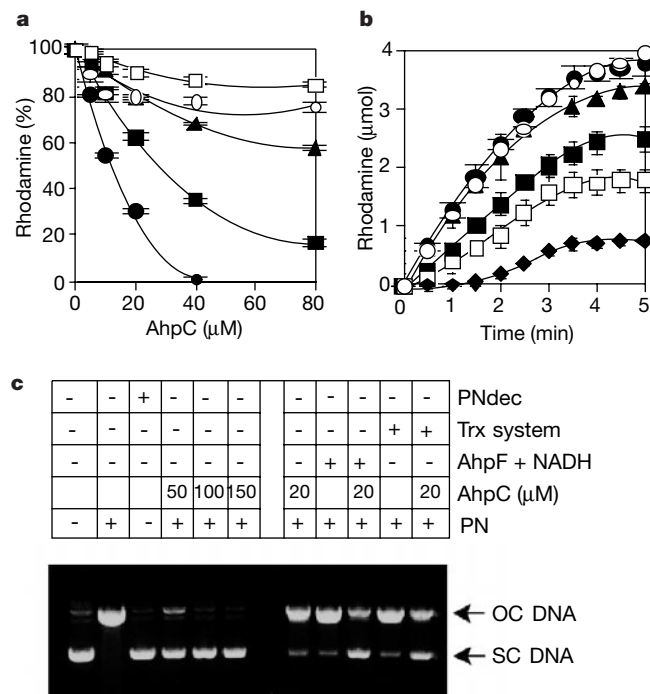


Figure 1 AhpC from *S. typhimurium* protects dihydrorhodamine and DNA from peroxynitrite-induced oxidation. **a**, Protection of dihydrorhodamine without catalytic turnover of AhpC. Reactions contained KPi buffer with 100 μM dihydrorhodamine and indicated concentrations of AhpC, either wild-type (filled circles), mutant C165A (filled squares), C46A (filled triangles), or C46/165A (open circles), or AhpC pre-oxidized with TBH (open squares). Peroxynitrite (20 μM) was added and rhodamine formation measured by absorbance at 500 nm wavelength. **b**, Protection of dihydrorhodamine during catalytic turnover of AhpC. Reactions as in **a** also contained 50 μM NADH and no protein (filled circles), 1 μM AhpC alone (open circles), 5 μM AhpF alone (filled triangles), or 5 μM AhpF plus AhpC at 1 μM (filled squares), 5 μM (open squares) or 10 μM (diamonds). Peroxynitrite was infused to achieve 8.4 μM min⁻¹ for 5 min. **c**, Protection of DNA. Reactions contained 1 μg of pBluescript II plasmid and indicated amounts of AhpC with or without 500 nM AhpF and 50 μM NADH, or 0.5 units of Trx, 2 μM Trx and 50 μM NADPH (Trx system) in KPi buffer pH 6.5. Peroxynitrite (PN) or decomposed peroxynitrite (PNdec) was infused to achieve 25 μM min⁻¹ for 3 min. Agarose gel electrophoresis separated supercoiled (SC) and open circular (nicked) (OC) DNA. All results are from ≥3 experiments (error bars show ± s.d.).

At a molar excess of AhpC over peroxynitrite, AhpC protected dihydrorhodamine completely against oxidation (Fig. 1a). Mutation of Cys 165 to Ala (C165A) diminished the protective effect of AhpC slightly. The mutant C46A was markedly less effective than wild type, and the double mutant C46/165A was scarcely protective. Likewise, protection was lost when the Cys residues in AhpC were oxidized by pre-exposure to *t*-butyl hydroperoxide (TBH) (Fig. 1a). Next we tested whether AhpC could act catalytically during perfusion with peroxynitrite in molar excess (Fig. 1b). Under these conditions, AhpC alone did not protect dihydrorhodamine unless we also provided pure AhpF and NADH to reduce AhpC. With the full system, AhpC displayed a concentration-dependent ability to protect dihydrorhodamine from a 40-fold molecular excess of peroxynitrite during 5 min of perfusion (Fig. 1b). DNA is a critical physiological target of peroxynitrite¹¹. At a molar excess of AhpC over peroxynitrite, AhpC alone protected supercoiled DNA from

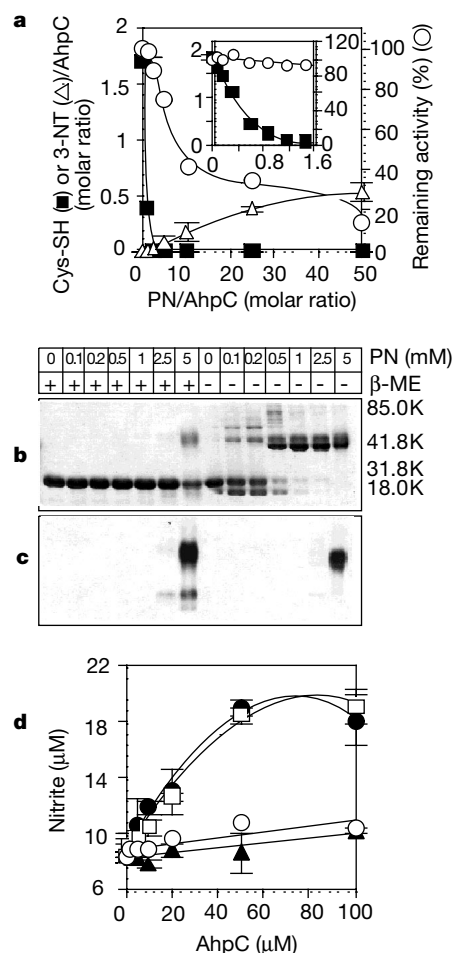


Figure 2 Catabolism of peroxynitrite by AhpC. **a**, Reversible oxidation of Cys residues. Samples of AhpC (100 μM) were pretreated with excess peroxynitrite (PN) as indicated, then assayed for TBH reductase activity (open circles), free Cys-SH content (filled squares), and 3-nitrotyrosine (3-NT) (open triangles). Inset, results at a low excess of peroxynitrite. **b**, Intermonomeric disulphide bond formation and cross-linking. AhpC treated with excess peroxynitrite as in **a** was boiled for 5 min in Laemmli buffer with or without β-mercapthoethanol (β-ME), separated by SDS-polyacrylamide gel electrophoresis and stained with Coomassie. Numbers on right, relative molecular mass *M_r*. **c**, Western blot with anti-nitrotyrosine monoclonal antibody (1 μg, Alexis) using a duplicate of the gel shown in **b**. **d**, Nitrite formation: dependence on Cys 46. Peroxynitrite (20 μM) was added to 100 μl of KPi buffer in the presence of indicated amounts of AhpC, either wild-type (filled circles) or mutants C46A (filled triangles), C165A (open squares) or C46/165A (open circles) for 15 min before measuring nitrite. All results are from ≥3 experiments (error bars show ± s.d.).

single-strand breaks. At a molar excess of peroxynitrite over AhpC, protection was evident only in the presence of AhpC + AhpF + NADH, or AhpC + thioredoxin reductase (TR) + thioredoxin (Trx) + NADPH (Fig. 1c).

To test if peroxynitrite reversibly oxidizes the Cys residues of AhpC, we pretreated AhpC with peroxynitrite at a 0.2- to 50-fold molar excess, measured the free Cys-SH groups remaining, and then tested whether the Cys-SH groups could be regenerated by AhpF + NADH as assessed by the ability of the pretreated AhpC to carry out the reduction of TBH (a Cys-dependent process¹²) (Fig. 2a). AhpC regained over 90% of its TBH reductase activity following exposure to a 2-fold molar excess of peroxynitrite, even though the Cys-SH content fell from 1.7 to 0 mol Cys-SH per mol AhpC monomer. Each molecule of peroxynitrite led to the oxidation of 2 Cys (Fig. 2a inset). In contrast, nitration of AhpC became detectable only when peroxynitrite was at ≥ 10 -fold molar excess over AhpC, and AhpC was not allowed to complete the catalytic cycle (Fig. 2a). Even at 50-fold molar excess of peroxynitrite over non-cycling AhpC, less than one of the five Tyr residues per AhpC monomer was detectably nitrated (Fig. 2b), a modification confirmed by immunoblot (Fig. 2c), and this was associated with irreversible crosslinking (Fig. 2b). Lower concentrations of peroxynitrite only caused disulphide bond formation (Fig. 2b). Thus, the first residues in AhpC to react with peroxynitrite are Cys, not Tyr, and the reaction leads to reversible disulphide bonding.

Spontaneous decomposition of peroxynitrite in pure solution leads to the accumulation of nitrate ($\sim 70\%$) and nitrite ($\sim 30\%$)¹. In contrast, wild-type AhpC and C165A caused the nearly quantitative conversion of peroxynitrite into nitrite, while C46A and C46/165A were ineffective (Fig. 2d). Thus, AhpC catabolizes peroxynitrite to nitrite; Cys 46 is required, but Cys 165 and a reducing system are not.

Peroxy-nitrite might react with AhpC Cys 46 via either oxidation or nitrosation. We sought to trap a putative sulphenic acid intermediate through its reaction with 7-chloro-4-nitrobenz-2-oxa-1,3-diazole (NBD-Cl), which will react with Cys-SH or Cys-SOH, but not with Cys-SO₂H, Cys-SO₃H, Cys-SNO or Cys-SS-Cys¹³. To preclude formation of Cys-SS-Cys, we used C165A. We anaerobically reacted C165A sequentially with peroxynitrite and NBD-Cl and characterized the products by electrospray ionization mass spectrometry (Fig. 3). Untreated C165A yielded a single peak within 1 atomic mass unit (a.m.u.) of the predicted mass, 20,581 (Fig. 3a), and was modified by NBD (164 a.m.u.) to produce a peak at 20,745 a.m.u. (Fig. 3b). After pre-treatment with subequimolar peroxynitrite, the major peak shifted to 20,761 a.m.u., demonstrating that one atom of oxygen was incorporated (Fig. 3c). Decomposed peroxynitrite caused no peak shift, ruling out effects from any residual H₂O₂ in the preparation. When C46A was reacted with peroxynitrite and NBD-Cl, the major peak remained at 20,745 a.m.u. (Fig. 3d), indicating that Cys 46 is the site of oxidation.

Stopped-flow spectroscopy allowed us to quantify the reactivity of AhpC with peroxynitrite. Spontaneous decomposition of

peroxynitrite at pH 7.0 and room temperature followed first-order kinetics with a rate constant of $0.51 \pm 0.04 \text{ s}^{-1}$. Scans at 10-ms intervals over the first 100 ms were nearly superimposable (Fig. 4a). In contrast, in the presence of AhpC, peroxynitrite was completely decomposed within 100 ms (Fig. 4b) with a second-order rate constant of $1.51 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ (Table 1). The reaction was slowed by 2.5–3 orders of magnitude if the Cys residues in AhpC were first oxidized with TBH or mutated (Table 1). As a further control, β -lactamase, an enzyme of similar mass containing three Cys residues, purified from the same bacteria, reacted with peroxynitrite 1,600 times more slowly than AhpC (Table 1).

Cys residues that serve as catalytic sites often have low pK_a values¹⁴. Consistent with this, the pH-dependence of the reactivity of C165A and C46A with a sulphhydryl-alkylating agent revealed that the pK_a for Cys 46 was < 5 . In contrast, the pK_a for Cys 165 was 8.7, close to that of free Cys. The pH optimum (6.72) for reaction of AhpC with peroxynitrite (Fig. 4d) is compatible with intracellular function.

Finally, AhpC homologues purified from two other phylogenetically distant pathogens, *M. tuberculosis* and *H. pylori*, reacted with peroxynitrite nearly as rapidly as AhpC from *S. typhimurium* (Fig. 4c; Table 1).

The remarkable rate of the reaction of peroxynitrite with AhpC, and the facility with which oxidized AhpC is reduced by AhpF or by the thioredoxin system, suggest that AhpC is likely to protect a cell from fluxes of peroxynitrite at sites where the intracellular concentration of AhpC is sufficiently high, and where the rate of peroxynitrite formation does not greatly exceed the reducing potential of a cell to replenish oxidized AhpC. The rates reported here for catabolism of peroxynitrite by bacterial peroxiredoxins are among the fastest known for peroxynitrite's reactions, despite the lack of haem and selenium. Cytochrome *c* oxidase, a haem protein, displayed peroxynitrite reductase activity ($10^6 \text{ M}^{-1} \text{ s}^{-1}$)¹⁵. Under aerobic conditions, this reaction would be outcompeted by O₂ ($10^8 \text{ M}^{-1} \text{ s}^{-1}$). Glutathione peroxidase, a selenoprotein, has the largest rate constant reported

Table 1 Second-order rate constants for peroxynitrite catabolism by bacterial peroxiredoxins and control proteins

Protein	Species	Pretreatment	Second-order rate constant ($\text{M}^{-1} \text{ s}^{-1}$)
AhpC wild type	<i>S. typhimurium</i>	None	$1.51 (\pm 0.04) \times 10^6$ *
AhpC wild type	<i>S. typhimurium</i>	Oxidized	$3.00 (\pm 0.06) \times 10^2$ †
AhpC C165A	<i>S. typhimurium</i>	None	$5.52 (\pm 0.19) \times 10^4$ †
AhpC C46A	<i>S. typhimurium</i>	None	$6.30 (\pm 0.05) \times 10^3$ †
AhpC C46/165A	<i>S. typhimurium</i>	None	$2.44 (\pm 0.04) \times 10^3$ †
AhpC wild type	<i>M. tuberculosis</i>	None	$1.33 (\pm 0.08) \times 10^6$ *
AhpC wild type	<i>H. pylori</i>	None	$1.21 (\pm 0.05) \times 10^6$ *
β -lactamase	<i>E. coli</i>	None	$< 2.0 \times 10^3$ ‡

* Means $\pm$ s.d. of three experiments with independent preparations of protein each tested at multiple concentrations.

† Means $\pm$ s.d. of >3 runs at each protein concentration.

‡ Mean from two experiments.

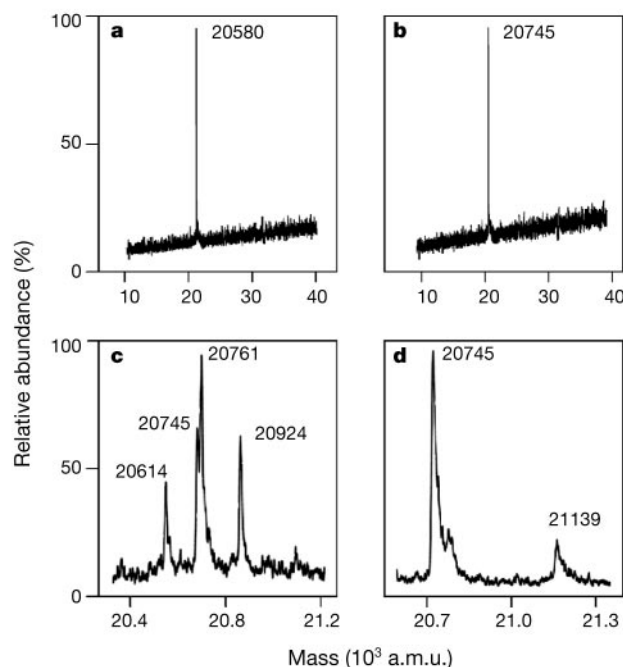


Figure 3 Sulphenic acid formation at Cys 46 in peroxynitrite-treated AhpC: identification by electrospray ionization mass spectrometry. **a**, AhpC C165A. **b**, NBD-modified AhpC C165A. **c**, NBD-modified AhpC C165A (100 μM) pretreated with peroxynitrite (75 μM). **d**, NBD-modified AhpC C46A pretreated with peroxynitrite as in **c**.

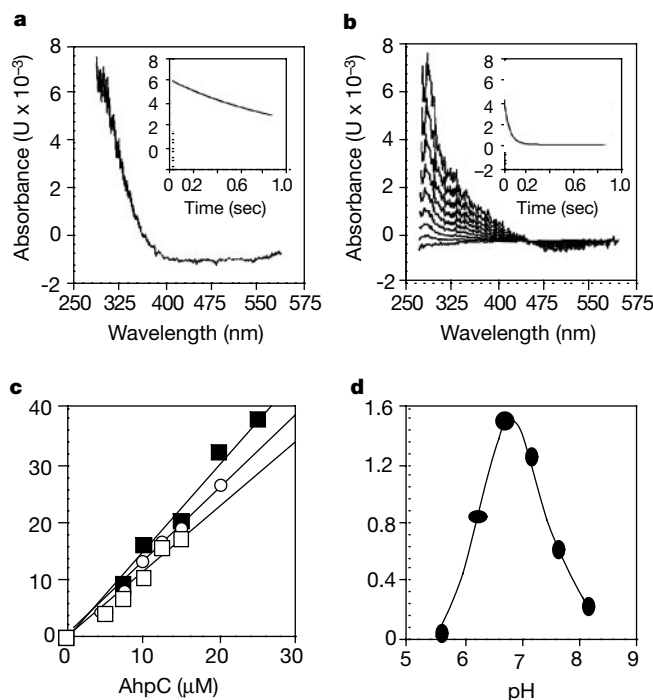


Figure 4 Kinetics of peroxynitrite catabolism by bacterial AhpC. **a**, Stopped-flow spectroscopy of peroxynitrite decomposition. Shown are nine overlying scans taken at 10-ms intervals after mixing. Inset, Time course monitored at 322 nm. **b**, Same as **a** except that AhpC (15 μ M) was injected. **c**, Determination of second-order rate constants. Experiments were performed in the presence of increasing concentrations of AhpC

for reaction of peroxynitrite with a physiological target (8×10^6 $\text{M}^{-1} \text{s}^{-1}$)¹⁶. However, many species lack a glutathione system, including mycobacteria. Widespread expression⁸ or overexpression¹⁷ of AhpC by clinical isolates of *M. tuberculosis* may thus help explain their resistance to peroxynitrite¹⁸.

The catalytic cycle of AhpC in reaction with alkylperoxide begins with formation of the cysteine sulphenic acid (Cys-OH) at Cys 46 (refs 13, 19). Disulphide bonding with Cys 165' in the opposite chain of the AhpC dimer is followed by NADH-dependent reduction of the disulphide by AhpF (ref. 12). Our experiments show that the reaction between AhpC and peroxynitrite proceeds similarly, beginning with nucleophilic attack of the Cys 46 thiolate on the O-O bond, resulting in the transfer of one oxygen atom to Cys 46 and release of nitrite. There has been little previous evidence that reactive nitrogen intermediates can oxidize protein-bound Cys residues to Cys-SOH (ref. 20) or that oxidation of a biological molecule by reactive nitrogen intermediates can be reversible and functionally consequential. Nor is there precedent for a selenium-free Cys residue reacting as rapidly with peroxynitrite as does Cys 46 of AhpC. The special nature of this reaction is emphasized by the lack of a comparable reaction with free Cys (ref. 21), the Cys in glutathione, the three Cys in β -lactamase, the two free Cys involved in the catalytic cycle of AhpF, or Cys 165 of AhpC. Features of the microenvironment that may facilitate the reactivity of Cys 46 with peroxynitrite include its low pKa and the ability of intermonomeric disulphide bond formation to protect nascent Cys-OH from irreversible oxidation to the sulphinic or sulphonic acids.

In contrast to AhpC's ready reaction with peroxynitrite, we were unable to demonstrate any reaction of AhpC with NO, NO₂⁻ or S-nitrosoglutathione (data not shown). Nonetheless, *ahpC* protected cells against the bactericidal and cytotoxic actions of NO₂⁻, S-nitrosoglutathione and the products of NOS2 (ref. 8). This seeming contradiction can be resolved by hypothesizing that diverse forms of reactive nitrogen intermediates achieve or enhance their cytotoxic potential after generating NO, which encounters O₂,

purified from *S. typhimurium* (filled squares), *M. tuberculosis* (open circles) or *H. pylori* (open squares). Apparent second-order rate constants were determined from slopes of the plots shown. **d**, pH optimum. Values for k_{obs} were determined in KPi buffer with final pH ranging from 5.5 to 8.0 and corrected for spontaneous decomposition in the absence of AhpC at each pH.

resulting in conversion into peroxynitrite.

Most studies of the reaction of peroxynitrite with proteins have focused on tyrosine nitration²², generally considered damaging and irreversible. However, a signalling role²³ is suggested by evidence that protein tyrosine nitration can be enzymatically reversed²⁴. The demonstration that peroxynitrite can reversibly oxidize a protein active site Cys strengthens the possibility that low levels of peroxynitrite may participate in signal cascades. As precedent, some growth factors elicit production of H₂O₂, which temporarily inactivates tyrosine phosphatases through reversible formation of Cys-SOH at the active site, permitting protein tyrosine phosphorylation²⁵.

We have shown here that three peroxiredoxins from phylogenetically divergent pathogens all react extremely rapidly with peroxynitrite. Reactivity is based on reversible oxidation of the N-terminal Cys, a residue conserved throughout the superfamily. Thus, the ability to act as a peroxynitrite reductase may be shared by many peroxiredoxins. Reciprocally, the utility of the peroxynitrite reductase reaction may help account for the wide distribution of peroxiredoxins. Thus the present observations may have relevance for the control of host-pathogen interactions, inflammation, mutagenesis, ageing and homeostasis. □

Methods

Peroxyntite was prepared by quenched-flow synthesis from acidified nitrite and hydrogen peroxide²⁶, and by biphasic synthesis from isoamyl nitrite and hydrogen peroxide²⁷ (product formation experiments; Fig. 2d).

Open reading frames (ORFs) of *ahpF*, *ahpC* and its mutants C46A, C165A and C46/165A from *S. typhimurium* were amplified by polymerase chain reaction (PCR) from pSty-CF (ref. 8). AhpC ORFs from *M. tuberculosis* and *H. pylori* were amplified by PCR from pPs-MtahpC (ref. 8) and ATCC NCTC 11637, respectively. ORFs were cloned into pET11c and expressed in *E. coli* BL21(DE3) (Novagen). AhpC from *S. typhimurium* and its mutants or AhpF were purified to homogeneity from isopropyl-1-thio- β -D-galactoside (IPTG)-treated cells by phenyl- (AhpC) or blue- (AhpF) sepharose and diethylaminoethyl (DEAE) chromatography²⁸. AhpC homologues from *M. tuberculosis* and *H. pylori* were purified to homogeneity by phenyl sepharose and MonoQ (Pharmacia) anion exchange or gel filtration and MonoQ chromatography²⁹ respectively.

For the determination of pKa values, AhpC C46A and C165A (20 μ M) were treated with iodoacetamide (200 μ M) in 100 mM potassium phosphate, 1 mM EDTA at pH values from 4 to 10 for 45 min at room temperature. DTNB (5,5'-dithiobis(2-nitrobenzoic acid); 200 μ M) was added, and sulphhydryl content determined by absorbance at a wavelength of 412 nm.

Stopped-flow experiments were performed on an OLIS RSM-16 instrument. One syringe contained 5 μ M peroxyxynitrite in 3 mM NaOH, and the other contained 100 mM potassium phosphate pH 7.0, 100 μ M DTPA (KPi buffer). Equal volumes were injected into a stopped-flow cell, and the pH of the reaction (6.75) was measured at the outlet. 400 scans were collected for each run, and k_{obs} was obtained by fitting experimental data to a single exponential decay function.

TBH reductase activity was measured in the presence of NADH (200 μ M), AhpF (500 nM) and AhpC (1 μ M), and expressed relative to untreated AhpC. Sulphydryl content was measured with DTNB. Tyrosine nitration was measured by absorbance at 430 nm.

Received 6 June; accepted 10 July 2000.

- Pfeiffer, S. *et al.* Metabolic fate of peroxyxynitrite in aqueous solution. Reaction with nitric oxide and pH-dependent decomposition to nitrite and oxygen in a 2:1 stoichiometry. *J. Biol. Chem.* **272**, 3465–3470 (1997).
- Coddington, J. W., Hurst, J. K. & Lyman, S. V. Hydroxyl radical formation during peroxyxynitrous acid decomposition. *J. Am. Chem. Soc.* **121**, 2438–2443 (1999).
- Merényi, G., Lind, J., Goldstein, S. & Czapski, G. Peroxyxynitrous acid homolyzes into $\cdot\text{OH}$ and $\cdot\text{NO}_2$ radicals. *Chem. Res. Toxicol.* **11**, 712–713 (1998).
- Niles, J. C., Burney, S., Singh, S. P., Wishnok, J. S. & Tannenbaum, S. R. Peroxyxynitrite reaction products of 3',5'-di-O-acetyl-8-oxo-7,8-dihydro-2'-deoxyguanosine. *Proc. Natl Acad. Sci. USA* **96**, 11729–11734 (1999).
- Wink, D. A. *et al.* Nitric oxide protects against cellular damage and cytotoxicity from reactive oxygen species. *Proc. Natl Acad. Sci. USA* **90**, 9813–9817 (1993).
- Jin, D.-Y. & Jeang, K.-T. in *Antioxidation and Redox Regulation of Genes* (eds Sen, C. K., Sies, H. & Baeuerle, P. A.) 381–407 (Academic, New York, 2000).
- Storz, G. *et al.* An alkyl hydroperoxide reductase induced by oxidative stress in *Salmonella typhimurium* and *Escherichia coli*: genetic characterization and cloning of *ahpC*. *J. Bacteriol.* **171**, 2049–2055 (1989).
- Chen, L., Xie, Q. & Nathan, C. Alkyl hydroperoxide reductase subunit C (AhpC) protects bacterial and human cells against reactive nitrogen intermediates. *Mol. Cell* **1**, 795–805 (1998).
- Nathan, C. & Shiloh, M. U. Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. *Proc. Natl Acad. Sci. USA* **97**, 8844–8848 (2000).
- Wilson, T., de Lisle, G. W., Marcinkeviciene, J. A., Blanchard, J. S. & Collins, D. M. Antisense RNA to *ahpC*, an oxidative stress defence gene involved in isoniazid resistance, indicates that AhpC of *Mycobacterium bovis* has virulence properties. *Microbiology* **144**, 2687–2695 (1998).
- Szabo, C., Zingarelli, B., O'Connor, M. & Salzman, A. L. DNA strand breakage, activation of poly (ADP-ribose) synthetase, and cellular energy depletion are involved in the cytotoxicity of macrophages and smooth muscle cells exposed to peroxyxynitrite. *Proc. Natl Acad. Sci. USA* **93**, 1753–1758 (1996).
- Ellis, H. R. & Poole, L. B. Roles for the two cysteine residues of AhpC in catalysis of peroxide reduction by alkyl hydroperoxide reductase from *Salmonella typhimurium*. *Biochemistry* **36**, 13349–13356 (1997).
- Ellis, H. R. & Poole, L. B. Novel application of 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole to identify cysteine sulfenic acid in the AhpC component of alkyl hydroperoxide reductase. *Biochemistry* **36**, 15013–15018 (1997).
- Claiborne, A. *et al.* Protein-sulfenic acids: diverse roles for an unlikely player in enzyme catalysis and redox regulation. *Biochemistry* **38**, 15407–15416 (1999).
- Briviba, K., Kissner, R., Koppenol, W. H. & Sies, H. Kinetic study of the reaction of glutathione peroxidase with peroxyxynitrite. *Chem. Res. Toxicol.* **11**, 1398–1401 (1998).
- Pearce, L. L., Pitt, B. R. & Peterson, J. The peroxyxynitrite reductase activity of cytochrome c oxidase involves a two-electron redox reaction at the heme a (3)-Cu(B) site. *J. Biol. Chem.* **274**, 35763–35767 (1999).
- Sherman, D. R. *et al.* Compensatory *ahpC* gene expression in isoniazid-resistant *Mycobacterium tuberculosis*. *Science* **272**, 1641–1643 (1996).
- Yu, K. *et al.* Toxicity of nitrogen oxides and related oxidants on mycobacteria: *M. tuberculosis* is resistant to peroxyxynitrite anion. *Tuber. Lung Dis.* **79**, 191–198 (1999).
- Choi, H.-J., Kang, S. W., Yang, C.-H., Rhee, S. G. & Ryu, S.-E. Crystal structure of a novel human peroxidase enzyme at 2.0 Å resolution. *Nature Struct. Biol.* **5**, 400–406 (1998).
- Becker, K., Savvides, S. N., Keese, M., Schirmer, R. H. & Karplus, P. A. Enzyme inactivation through sulphydryl oxidation by physiologic NO-carriers. *Nature Struct. Biol.* **5**, 267–271 (1998).
- Radi, R., Beckman, J. S., Bush, K. M. & Freeman, B. A. Peroxyxynitrite oxidation of sulphydryls. The cytotoxic potential of superoxide and nitric oxide. *J. Biol. Chem.* **266**, 4244–4250 (1991).
- Beckman, J. S. & Koppenol, W. H. Nitric oxide, superoxide, and peroxyxynitrite: the good, the bad, and ugly. *Am. J. Physiol.* **271**, C1424–C1437 (1996).
- Go, Y. M. *et al.* Evidence for peroxyxynitrite as a signaling molecule in flow-dependent activation of c-Jun NH(2)-terminal kinase. *Am. J. Physiol.* **277**, H1647–H1653 (1999).
- Kamisaki, Y. *et al.* An activity in rat tissues that modifies nitrotyrosine-containing proteins. *Proc. Natl Acad. Sci. USA* **95**, 11584–11589 (1998).
- Denu, J. M. & Tanner, K. G. Specific and reversible inactivation of protein tyrosine phosphatases by hydrogen peroxide: evidence for a sulfenic acid intermediate and implications for redox regulation. *Biochemistry* **37**, 5633–5642 (1998).
- Koppenol, W. H., Kissner, R. & Beckman, J. S. Synthesis of peroxyxynitrite: to go with the flow or on solid grounds? *Methods Enzymol.* **269**, 296–302 (1996).
- Uppu, R. M. & Pryor, W. A. Biphasic synthesis of high concentrations of peroxyxynitrite using water-insoluble alkyl nitrite and hydrogen peroxide. *Methods Enzymol.* **269**, 322–329 (1996).
- Poole, L. B. & Ellis, H. R. Flavin-dependent alkyl hydroperoxide reductase from *Salmonella typhimurium*. 1. Purification and enzymatic activities of overexpressed AhpF and AhpC proteins. *Biochemistry* **35**, 56–64 (1996).
- O'Toole, P. W., Logan, S. M., Kostrzynska, M., Wadstrom, T. & Trust, T. J. Isolation and biochemical and molecular analyses of a species-specific protein antigen from the gastric pathogen *Helicobacter pylori*. *J. Bacteriol.* **173**, 505–513 (1991).

Acknowledgements

We thank L. Chen for generating *ahpC* mutations, G. St John for the *H. pylori* clone, H. Erdjument-Bromage and P. Tempst for protein sequencing and T. Sakmar for access to his stopped-flow spectrophotometer. This work was supported by a Norman and Rosita Winston fellowship (R.B.) and by an NIH grant (C.N.).

Correspondence and requests for materials should be addressed to C.N. (e-mail: cnathan@med.cornell.edu).

Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase

John W. Trauger*, Rahul M. Kohli*, Henning D. Mootz†, Mohamed A. Marahiel† & Christopher T. Walsh*

* Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA

† Biochemie/Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Strasse, 35032 Marburg, Germany

In the biosynthesis of many macrocyclic natural products by multidomain megasynthases, a carboxy-terminal thioesterase (TE) domain is involved in cyclization and product release^{1,2}; however, it has not been determined whether TE domains can catalyse macrocyclization (and elongation in the case of symmetric cyclic peptides) independently of upstream domains. The inability to decouple the TE cyclization step from earlier chain assembly steps has precluded determination of TE substrate specificity, which is important for the engineered biosynthesis of new compounds¹. Here we report that the excised TE domain from tyrocidine synthetase efficiently catalyses cyclization of a decapeptide-thioester to form the antibiotic tyrocidine A, and can catalyse pentapeptide-thioester dimerization followed by cyclization to form the antibiotic gramicidin S. By systematically varying the decapeptide-thioester substrate and comparing cyclization rates, we also show that only two residues (one near each end of the decapeptide) are critical for cyclization. This specificity profile indicates that the tyrocidine synthetase TE, and by analogy many other TE domains, will be able to cyclize and release a broad range of new substrates and products produced by engineered enzymatic assembly lines.

An enormous range of medically important polyketide and peptide natural products assembled by modular polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs) and mixed PKS/NRPS systems have macrocyclic structures, including the antibiotics erythromycin (PKS) and daptomycin (NRPS), the immunosuppressants cyclosporin (NRPS) and rapamycin (PKS/NRPS), and the antitumour agent epothilone (PKS/NRPS). PKSs and NRPSs are large, multifunctional proteins that are organized into sets of functional domains termed modules^{1,2}. The order of modules corresponds directly to the sequence of monomers in the product. Synthetic intermediates are covalently tethered by thioester linkages to a carrier protein domain in each module. The thiol tether on each carrier domain is phosphopantetheine, which is attached to a conserved serine residue in the carrier protein in a post-translational priming reaction catalysed by a phosphopantetheinyltransferase³. Chain initiation involves loading a specific monomer onto the thiol tether of each carrier protein. Subsequent chain elongation steps involve transfer of the growing chain from an upstream carrier protein to the adjacent downstream carrier-protein-bound monomer. Release of the full-length chain

from the most C-terminal carrier domain is almost always catalysed by a C-terminal TE domain of relative molecular mass 28,000–35,000 (M_r 28K–35K)¹. Thioesterase domains of PKS and NRPS use a catalytic triad consisting of a serine residue (or cysteine in some cases), histidine and an acidic residue^{4,5}. Thioester cleavage involves sequential acylation and deacylation of the active-site serine^{5–7}. In the acylation step, the full-length chain is transferred from the thiol tether of the upstream carrier protein to the serine residue. Deacylation of the resulting acyl-O-TE intermediate occurs either by intramolecular cyclization to form macrolactones or macrolactams or by hydrolysis.

It was not clear at the outset whether an isolated TE domain would retain cyclization activity; when the deoxyerythronlide B synthase TE domain was overproduced and purified it had hydrolysis activity, but did not cyclize synthetic acyl-thioester substrates^{7,8}. The absence of cyclization in these cases might reflect that either the simplified substrate analogues lacked functional groups required for substrate recognition or the TE cannot independently catalyse cyclization. With regard to TE specificity, engineered PKS and NRPS systems in which domains have been added, substituted, or deleted can produce new polyketides^{9–14} and peptides^{15–17}. For example, engineered variants of deoxyerythronlide B synthase, which normally produces a 14-membered macrolactone, produce a variety of cyclic products ranging in size from 6- to 16-membered rings^{9–13}; however, the product yield of an engineered synthase is often much lower than that of its natural counterpart¹². This limitation may reflect the specificity of individual domains within a synthase. For example, analysis of a peptide-bond-forming C domain from an NRPS revealed that this domain has low selectivity for the upstream residue, but high selectivity for the downstream residue¹⁸. As a synthase engineered to assemble a new polyketide or peptide chain will be useful only if the TE can efficiently release the product by cyclization or hydrolysis, characterizing TE specificity is a critical step toward robust engineered synthases.

We wanted to determine whether the TE domain from the tyrocidine NRPS (Fig. 1a), which catalyses assembly of the cyclic decapeptide antibiotic tyrocidine A^{19,20}, can independently catalyse peptide cyclization. We replaced full-length TycC (M_r 724K) with overexpressed and purified TycC TE domain (M_r 28K) and replaced

decapeptide-S-PCP (the natural substrate of the TE domain) with a synthetic peptide *N*-acetylcysteamine (NAC) thioester (peptide-SNAC) (Fig. 1b). *N*-acetylcysteamine is structurally identical to the terminal portion of phosphopantetheine and thus a good mimic of the natural substrate decapeptide-S-PCP^{7,8}. When the decapeptide-SNAC corresponding to the tyrocidine A sequence

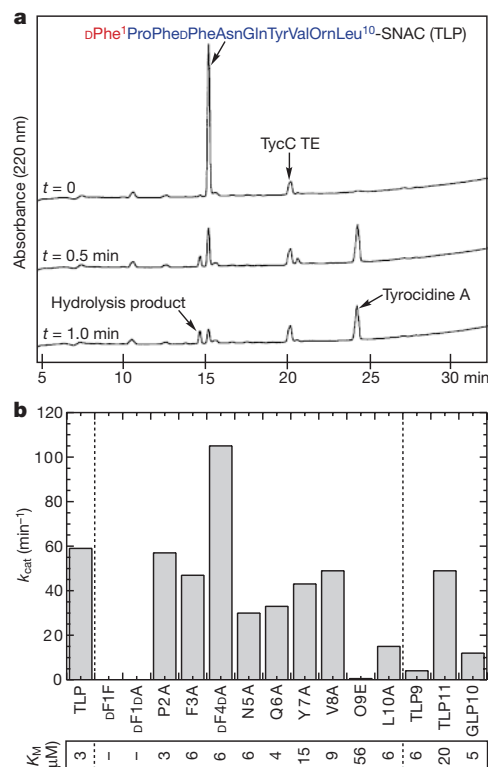


Figure 2 Cyclization activity and substrate specificity of the TycC TE domain. **a**, HPLC analysis of reactions that initially contained 2 μM TLP, 50 nM TycC TE and 25 mM MOPS (pH 7.0, 24 °C). **b**, Kinetic parameters for peptide-SNAC cyclization catalysed by the TE domain of TycC.

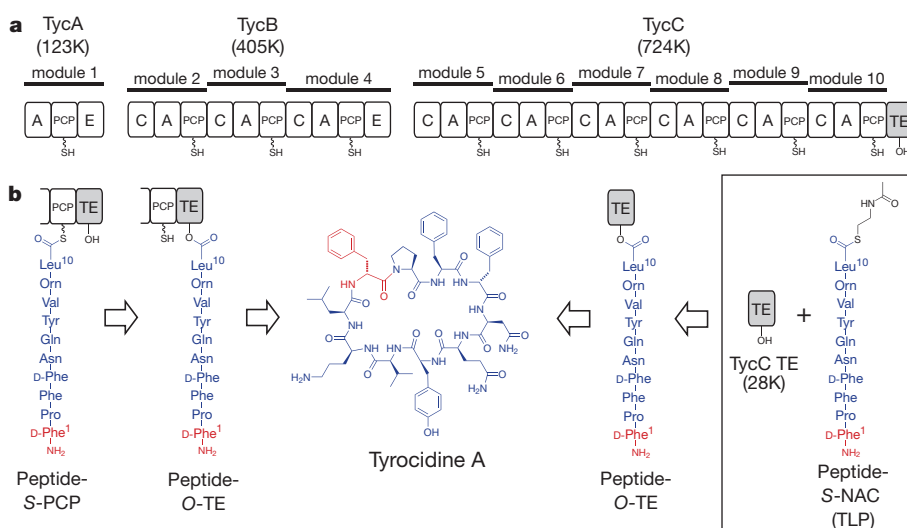


Figure 1 Role of the thioesterase (TE) domain in biosynthesis of the cyclic decapeptide antibiotic tyrocidine A. **a**, The tyrocidine non-ribosomal peptide synthetase from *B. brevis*. Synthetase subunits TycA, TycB and TycC are represented by a series of boxes. Each box represents a functional domain: A, adenylation (catalyses amino-acid activation); PCP, peptidyl carrier protein; C, condensation (catalyses peptide-bond formation); E, epimerization. The TE domain is shaded. Thiol (SH) and hydroxyl (OH) groups represent

phosphopantetheine and the TE active-site serine residue, respectively. **b**, Left, proposed mechanism of TE-domain catalysed macrocyclization and product release. Decapeptide intermediates are shown with the N-terminal residue d-Phe¹ highlighted in red. Right, Experimental system for probing the mechanism and specificity of TE-domain catalysed cyclization.

(D-Phe-Pro-Phe-D-Phe-Asn-Gln-Tyr-Val-Orn-Leu-SNAC (TLP); where Orn is ornithine) was incubated with purified TycC TE, efficient cyclization to tyrocidine A occurred as well as a minor flux of hydrolysis to the decapeptide (ratio of cyclization:hydrolysis, 6:1) (Fig. 2a). The cyclic product was identified as tyrocidine A by its co-elution with authentic tyrocidine A during high performance liquid chromatography (HPLC) and by mass spectrometry. Kinetic analysis of the cyclization reaction established a k_{cat} of 59 turnovers per minute and a K_M of 3 μM . No hydrolysis or cyclization was detectable under the reaction conditions in the absence of enzyme. Thus, the isolated TE domain retains a robust capacity to recognize the tyrocidine decapeptide chain and to catalyse efficiently cyclization to a 30-membered macrocycle. To our knowledge, the small flux to hydrolysis is not observed with the natural synthetase, and may reflect the absence of upstream domains. We measured background hydrolysis and cyclization rates for TLP, which revealed that the TE domain of TycC accelerates hydrolysis and cyclization by factors of 1×10^6 and 2×10^7 , respectively.

To explore the substrate specificity of TycC TE, we synthesized a series of mutant peptide-SNAC substrates that differ from the wild-type tyrocidine A sequence by a single amino-acid substitution. Specifically, we changed the amino-terminal residue D-Phe1 to either L-Phe or D-Ala, D-Phe4 to D-Ala, Orn9 to Glu (Orn was not changed to Ala because the resulting cyclic molecule would be unchanged and have poor water solubility), and each of the other seven residues to Ala. We determined kinetic parameters for the cyclization of each of the mutant substrates (Fig. 2b). Notably, mutation of the N-terminal residue D-Phe1 to either L-Phe (DF1F) or D-Ala (DF1DA) abolishes cyclization activity, indicating that recognition of both the stereochemistry and side chain of this residue is essential for cyclization. Hydrolysis of DF1F and DF1DA is observed, with kinetic parameters similar to the wild-type substrate TLP, indicating that these mutations affect the cyclization step and not peptide-O-TE formation. Recognition of Orn9 is also critical for cyclization: when changed to Glu (O9E), cyclization still occurs, but with k_{cat} decreased 100-fold and K_M increased 20-fold. Changing Orn to Glu affects cyclization and hydrolysis equally, indicating that the mutation affects the peptide-O-TE formation step. Mutants in which each of the remaining eight residues are changed to alanine (without changing L- or D- configuration) have relatively little effect on cyclization kinetics: k_{cat} values for all of these substrates are within a factor of 2 of the wild-type substrate TLP except for L10A (fourfold reduction in k_{cat}), and all of the K_M values are within a factor of 2 of TLP except Y7A (fivefold increase in K_M) and V8A (threefold increase in K_M). These results suggest that Tyr7, Val8 and Leu10 contribute to substrate recognition, although their contribution is much less than that of D-Phe1 and Orn9.

The observed specificity of the TE domain of TycC suggests that

various substrates that retain the key 'recognition residues' will be cyclized. To test this prediction, we synthesized 9-residue (D-Phe-Pro-Phe-Asn-Gln-Tyr-Val-Orn-Leu-SNAC; TLP9) and 11-residue (D-Phe-Pro-Phe-D-Phe-Asn-Ala-Gln-Tyr-Val-Orn-Leu-SNAC; TLP11) substrates in which one residue near the centre of the TLP sequence is either deleted or inserted. Both 9- and 11-membered substrates are cyclized by the TE domain of TycC (Fig. 2b). The 15-fold reduction in k_{cat} for TLP9 may result from strain in the cyclic conformation. These results show that TycC TE can catalyse formation of cyclic peptides with various ring sizes.

We next investigated whether the TycC TE could catalyse assembly of the cyclic decapeptide antibiotic gramicidin S. The amino-acid sequence of gramicidin S is a pentapeptide repeat, and the gramicidin S NRPS contains only five amino-acid activation modules (Fig. 3a), suggesting that the gramicidin TE domain (GrsB TE) catalyses dimerization of two pentapeptides and cyclization of the resulting decapeptide^{21,22}. A proposed mechanism⁶ for this process is shown in Fig. 3b. Comparison of the sequence of the tyrocidine decapeptide precursor with that of the gramicidin S pentapeptide precursor (D-Phe-Pro-Val-Orn-Leu) reveals that both peptides have the same two N-terminal residues (D-Phe-Pro) and the same three C-terminal residues (Val-Orn-Leu).

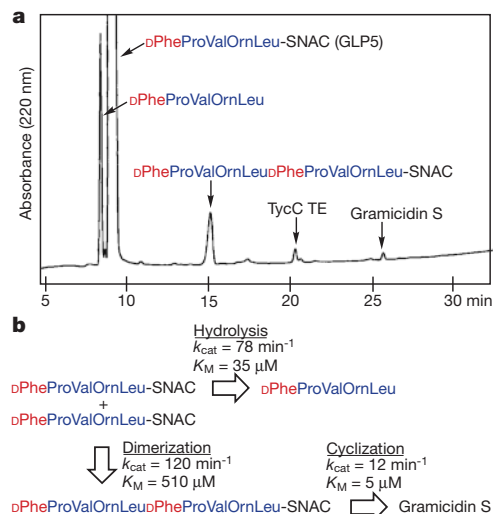


Figure 4 The tyrocidine synthetase thioesterase domain (TycC TE) catalyses dimerization of the pentapeptide-SNAC GLP5 and cyclization of the resulting decapeptide-SNAC to form gramicidin S. **a**, HPLC analysis of a reaction that initially contained 200 μM GLP5, 100 nM TycC TE and 25 mM MOPS (pH 7.0, 24 °C, 1 min reaction time). **b**, Kinetic parameters for GLP5 dimerization and hydrolysis and GLP10 cyclization catalysed by the TE domain of TycC.

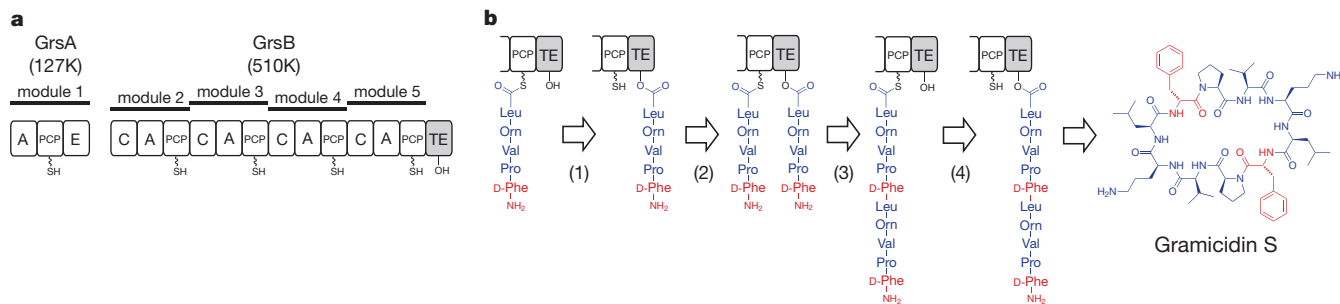


Figure 3 Role of the TE domain in biosynthesis of the cyclic decapeptide antibiotic gramicidin S. **a**, Gramicidin S synthetase from *B. brevis*. **b**, Proposed mechanism of pentapeptide dimerization and decapeptide cyclization catalysed by the TE domain of GrsB: (1) a pentapeptide is built up by the synthetase and transferred to the TE active-site

serine; (2) a second pentapeptide is built up; (3) the N-terminal amine of the pentapeptide-S-PCP reacts with the peptide-O-TE to form a decapeptide-S-PCP intermediate; and (4) the PCP-tethered decapeptide is transferred to the TE serine and cyclized.

As our results with TycC TE indicated that these common N- and C-terminal sequences are sufficient for substrate recognition, we reasoned that the TycC TE should be able to dimerize the gramicidin S pentapeptide-SNAC (D-Phe-Pro-Val-Orn-Leu-SNAC; GLP5) to form a decapeptide-SNAC that would then cyclize to gramicidin S. On incubation of GLP5 with TycC TE, we observed efficient chain dimerization and subsequent cyclization, as well as hydrolysis to the pentapeptide (Fig. 4). The identities of the products were confirmed by HPLC co-elution with authentic standards and by mass spectrometry. This result shows that the isolated TycC TE domain can catalyse peptide elongation and reveals the mechanism for pentapeptide dimerization and cyclization by gramicidin S synthetase, as well as the likely mechanism for assembly of other symmetric cyclic peptides such as the antitumour antibiotic thiocoraline²³. This finding complements our previous demonstration that the TE-containing protein EntF from enterobactin synthetase can catalyse two elongation cycles followed by cyclization to form enterobactin, a 12-membered trilactone⁶. Finally, the fact that the gramicidin S precursor decapeptide-SNAC (D-Phe-Pro-Val-Orn-Leu-D-Phe-Pro-Val-Orn-Leu-SNAC; GLP10), which differs from the tyrocidine A precursor TLP by the substitution of five central residues, cyclizes at a rate similar to TLP (fivefold lower k_{cat} and comparable K_M for GLP10 compared with TLP) further shows that the TycC TE can cyclize various substrates provided that the necessary 'recognition residues' near each end of the substrate are present. We note that ligation²⁴ and cyclization of synthetic acyl-thioesters may be a generally useful application of excised TE domains from PKS and NRPS systems. □

Methods

Peptide-SNAC synthesis

Peptides were prepared by automated solid-phase synthesis (0.3 mmol scale, diisopropylcarbodiimide/hydroxybenzotriazole (DIPCDI/HOBt) activation) on 2-chlorotrityl resin derivatized with the appropriate C-terminal amino acid using Fmoc-protected monomers (side-chain protecting groups used were: trityl for Asn and Gln, *t*-butyl for Tyr, and Boc for Orn) except for the N-terminal monomer, which was Boc-protected. The peptide was cleaved from the resin using 1:1:3 acetic acid:trifluoroethanol:dichloromethane (DCM) (3 h, 24 °C), then precipitated with *n*-hexane and the solvent removed by rotary evaporation. The protected peptide (1 eq.) was dissolved in tetrahydrofuran (THF) or dimethylformamide (DMF). A solution of DCC (1.2 eq.) and HOBt (1.2 eq.) in THF (or DMF) and *N*-acetylcysteine (2.5 eq.) were added, and the reaction stirred for 35 min at 24 °C. Potassium carbonate (0.6 eq.) was then added, and the reaction stirred for 3 h at 24 °C, filtered and concentrated. The protected peptide-SNAC was treated with 16:3:1 trifluoroacetic acid (TFA):DCM:*N*-acetylcysteine (3 h, 24 °C) and precipitated with ether. Reverse-phase (C_{18}) HPLC purification (20 to 50% acetonitrile in 0.1% TFA/water over 30 min) afforded the peptide-SNAC TFA salt (10–25% yield from the protected peptide) in more than 95% purity (by analytical HPLC) as a white solid. The identities of the all peptide-SNACs were verified by MALDI-TOF mass spectrometry.

Protein expression

TycC TE DNA was amplified from *Bacillus brevis* (ATCC 8185) chromosomal DNA using the primers 5'-ATAAGATCTCATAGCGCTTTGAGAGCAG-3' and 5'-ATAGGATCC TTTTCAGGATGAACAGTCTCTTG-3' and Vent DNA Polymerase (New England Biolabs). The fragment was digested with *Bam*HI and *Bgl*II and cloned into *Bgl*II-digested pQE60 (Qiagen), which adds the sequence GSRSHHHHHH to the C terminus of the expressed protein. Expression of the resulting plasmid pQE60-TE(TycC) in *Escherichia coli* M15[pREP4] (Qiagen) followed by Ni-NTA affinity chromatography yielded TycC TE (25 mg l⁻¹; > 95% pure by SDS-PAGE).

Assays

We carried out reactions in 25 mM MOPS, pH 7.0, in a total volume of 400 µl. Reactions were initiated by addition of TycC TE and quenched at various time points by the addition of 25 µl 1.7% TFA/water, flash frozen in liquid nitrogen and stored at -80 °C (for O9E, reactions were quenched by the addition of sodium phosphate, pH 5.3 to 100 mM). We then thawed the reactions, added 85 µl acetonitrile, and analysed them by analytical HPLC with monitoring at 220 nm (20–80% acetonitrile in 0.1% TFA/water (or in 25 mM potassium phosphate, pH 5.3, for O9E) over 35 min, Vydac protein and peptide C_{18} column). Initial rates were calculated using 1-min time points. Peptide-SNAC and reaction product concentrations were determined for all Tyr-containing peptides based on the estimated extinction coefficient ϵ (280 nm) = 1,280 M⁻¹ cm⁻¹, which agrees with the experimentally determined ϵ (280 nm) of TLP. For peptide-SNACs not containing Tyr, ϵ (220 nm) was determined experimentally, and concentrations of corresponding cyclic products determined by assuming equal ϵ (220 nm) values for the peptide-SNAC and cyclic product.

Product characterization

All cyclic products were characterized by MALDI-TOF mass spectrometry. Cyclic products enzymatically synthesized from TLP, P2A, L10A, TLP9, TLP11 and GLP10 were further characterized by ESI-ion trap mass spectrometry. Enzymatically synthesized (from TLP) and authentic tyrocidine A gave identical fragment ions, including four internal fragment ions (observed both with and without loss of NH₃ from Asn or Gln) that contain the Leu10-D-Phe1 dipeptide segment formed by head-to-tail cyclization, and at least two of the corresponding fragment ions were identified for cyclic peptides from P2A, L10A, TLP9 and TLP11, confirming that these products result from head-to-tail cyclization. For example, an Orn9 to Tyr7 ion was observed for each cycle (TLP cycle, M+H for O LFPFFNQY: calculated 1157.6, observed 1157.6; P2A cycle, M+H for OLAFFNQY: calculated 1131.6, observed 1131.5; L10A cycle, M+H for OAFPFNQY: calculated 1115.5, observed 1115.5; TLP9 cycle, M+H for OLPFFNQY: calculated 1009.5, observed 1009.3; TLP11 cycle, M+H for OLPFFNQY calculated 1228.6, observed 1228.6; the underline highlights the LF bond formed by head to tail condensation). Similarly, the GLP10 cycle and authentic gramicidin S gave the same fragmentation pattern, and one ion confirming head-to-tail cyclization was detected in both samples (GLP10 cycle, M+H for LFPVO LFPV: calculated 914.6, observed 914.5).

Received 23 May; accepted 3 July 2000.

- Cane, D. E., Walsh, C. T. & Khosla, C. Harnessing the biosynthetic code: combinations, permutations and mutations. *Science* **282**, 63–68 (1998).
- Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**, 2651–2674 (1997).
- Lambalot, R. H. *et al.* A new enzyme superfamily—the phosphopantetheinyl-transferases. *Chem. Biol.* **3**, 923–936 (1996).
- Lawson, D. M. *et al.* Structure of a myristoyl-ACP-specific thioesterase from *Vibrio harveyi*. *Biochemistry* **33**, 9382–9388 (1994).
- Li, J., Zittner, R., Derewenda, Z. S. & Meighen, E. A. Conversion of serine-114 to cysteine-114 and the role of the active site nucleophile in acyl transfer by myristoyl-ACP thioesterase from *Vibrio harveyi*. *Biochemistry* **35**, 9967–9973 (1996).
- Shaw-Reid, C. A. *et al.* Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E. coli* EntF catalyses both elongation and cyclolactonization. *Chem. Biol.* **6**, 385–400 (1999).
- Aggarwal, R., Caffrey, P., Leadlay, P. F., Smith, C. J. & Staunton, J. The thioesterase domain of the erythromycin-producing polyketide synthase: mechanistic studies *in vitro* to investigate its mode of action and substrate specificity. *J. Chem. Soc. Chem. Comm.* **15**, 1519–1520 (1995).
- Gokhale, R. S., Hunziker, D., Cane, D. E. & Khosla, C. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem. Biol.* **6**, 117–125 (1998).
- Cortes, J. *et al.* Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* **268**, 1487–1489 (1995).
- Kao, C. M. *et al.* Gain of function mutagenesis of the erythromycin polyketide synthase. 2. Engineered biosynthesis of an eight-membered ring tetraketide lactone. *J. Am. Chem. Soc.* **119**, 11339–11340 (1997).
- Jacobsen, J. R., Hutchinson, C. R., Cane, D. E. & Khosla, C. Precursor-directed biosynthesis of erythromycin analogs by an engineered polyketide synthase. *Science* **277**, 367–369 (1997).
- McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel 'unnatural' natural products. *Proc. Natl Acad. Sci. USA* **96**, 1846–1851 (1999).
- Tang, L., Fu, H. & McDaniel, R. Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases. *Chem. Biol.* **7**, 77–84 (2000).
- Xue, Y. & Sherman, D. H. Alternative modular polyketide synthase expression controls macrolactone structure. *Nature* **403**, 571–575 (2000).
- Stachelhaus, T., Schneider, A. & Marahiel, M. A. Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science* **269**, 69–72 (1995).
- de Ferra, F., Rodriguez, F., Tortora, O., Tosi, C. & Grandi, G. Engineering of peptide synthetases. *J. Biol. Chem.* **272**, 25304–25309 (1997).
- Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Construction of hybrid peptide synthetases by module and domain fusions. *Proc. Natl Acad. Sci. USA* **97**, 5848–5853 (2000).
- Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–489 (1999).
- Mootz, H. D. & Marahiel, M. A. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843–6850 (1997).
- Roskoski, R., Kleinkauf, H., Gevers, W. & Lipmann, F. Isolation of enzyme-bound peptide intermediates in tyrocidine biosynthesis. *Biochemistry* **9**, 4846–4851 (1970).
- Hori, K. *et al.* Molecular cloning and nucleotide sequence of the gramicidin S synthetase 1 gene. *J. Biochem. (Tokyo)* **106**, 639–645 (1989).
- Turgay, K., Krause, M. & Marahiel, M. A. Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol. Microbiol.* **6**, 529–546 (1992).
- Boger, D. L. & Ichikawa, S. Total syntheses of thiocoraline and BE-22179: establishment of relative and absolute stereochemistry. *J. Am. Chem. Soc.* **122**, 2956–2957 (2000).
- Jackson, D. Y. *et al.* A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science* **266**, 243–247 (1994).

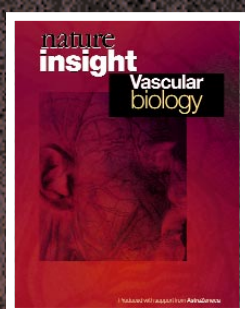
Acknowledgements

This work was supported by grants from the NIH (to C.T.W.), the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie (to M.A.M.), a NIH postdoctoral fellowship (to J.W.T.) and a PhD fellowship of the Stiftung Stipendien-Fonds des Verbandes der Chemischen Industrie (to H.D.M.).

Correspondence and requests for materials should be addressed to C.T.W. (e-mail: walsh@walsh.med.harvard.edu).

nature insight

Vascular biology



Research in vascular biology has boomed in recent years. The advances made have led to significant insights into the treatment not only of pulmonary vascular diseases such as atherosclerosis, ischaemic and congenital heart disorders, stroke, thrombosis and hypertension, but also diabetes and tumour development. But cardiovascular diseases remain a major cause of mortality worldwide, regardless of the recent advances in medical and surgical treatment. Indeed, as life expectancy in the developed world increases, cardiovascular conditions affecting the elderly are also likely to rise. And this escalation in cardiovascular disorders may not be confined to the developed world as recent data suggest that heart disease is also increasing among Asians and Chinese. As we are catapulted into the genomics era, exciting new therapeutic avenues are uncovered which may eventually lead to the tailoring of therapies to a patient's specific metabolic and genetic profile. To accelerate progress into the medical treatment of these disorders we must start by expanding our knowledge in the basic regulatory mechanisms underlying vascular biology. With this view in mind, this month's *Nature Insight* reveals the current research developments that are relevant to the understanding of the complex nature of vascular biology.

On page 221 Deepak Srivastava and Eric Olson provide an overview into the underlying genetics controlling cardiac development. The study of complex heart diseases has always been a difficult task, and in his review on page 227 Kenneth Chien discusses how the availability of genomic databases has empowered our understanding of these complex disorders. The atherosclerotic lesion has been recognized for many centuries and is the leading cause of death in the western world, often acting as the underlying trigger of heart attacks. On page 233 Aldons Jake Lusis provides insights into its complex aetiology and how therapeutic development may be achieved through genetic dissection of this progressive condition. Blood vessels have a more complex role in body homeostasis than merely acting as a simple conduit for the distribution of oxygen and nutrients. A background into the developmental process of angiogenesis is revealed by George Yancopoulos and colleagues on page 242, while on page 249 Peter Carmeliet and Rakesh Jain examine how this process is derailed in disease, in particular in tumorigenesis. Thrombin and platelets are important in myocardial infarction and other disorders. On page 258 Shaun Coughlin explains how our understanding of thrombin signalling may lead to new therapies. Finally, on page 265 Edward Rubin and Alan Tall highlight the promises of genomic technology in vascular biology.

We are pleased to acknowledge the financial support of AstraZeneca in producing this Insight. Of course, *Nature* carries the sole responsibility for all editorial content and peer-review. *Nature* has been revered in publishing many key studies in the field of vascular biology, we hope that you will find this collection of reviews enlightening and inspiring.

Andrea Kauffmann-Zeh Senior Editor

Ritu Dhand Insight Programme Editor

Publisher and liaison for corporate support **Liz Allen** (e.allen@nature.com)

A genetic blueprint for cardiac development

Deepak Srivastava*† & Eric N. Olson*

*Department of Molecular Biology and †Department of Pediatrics, University of Texas, Southwestern Medical Center at Dallas, 6000 Harry Hines Blvd, Dallas, Texas 75390-9148, USA (e-mail: dsriva@mednet.swmed.edu; eolson@hamon.swmed.edu)

Congenital heart disease is the leading non-infectious cause of death in children. It is becoming increasingly clear that many cardiac abnormalities once thought to have multifactorial aetiologies are attributable to mutations in developmental control genes. The consequences of these mutations can be manifest at birth as life-threatening cardiac malformations or later as more subtle cardiac abnormalities. Understanding the genetic underpinnings of cardiac development has important implications not only for understanding congenital heart disease, but also for the possibility of cardiac repair through genetic reprogramming of non-cardiac cells to a cardiogenic fate.

Formation of the heart involves a wondrous and precisely orchestrated series of molecular and morphogenetic events, and even subtle perturbation of this process can have catastrophic consequences in the form of congenital heart disease (CHD). The course of heart formation in the embryo has been known for decades but, until recently, we have known little about the genes that control its developmental programme. Insight has come from studies in vertebrate and invertebrate model organisms that have identified cardiac developmental control genes, which can now be investigated as possible culprits in human CHD. Of equal importance has been the mapping and identification of human CHD genes (see Box 1, Table 1). Insights into the structure–function relationships of the encoded proteins *in vivo* have been gained by correlation of particular amino-acid mutations with the associated cardiac malformations. Most genetic studies in model organisms involve homozygous loss-of-function mutations, which frequently result in severe phenotypes and are lethal in early embryogenesis, and so this approach identifies essential genes that act at critical steps in cardiac development. CHD in humans is, however, primarily a disease of haploinsufficiency, often involving more subtle phenotypes with variable penetrance, which are evident at birth or later. Thus, the two approaches yield different but complementary insights.

Formation of an organ as complex as the heart, with its many integrated structures and cell types, must involve a myriad of genes, many (or most) of which are not cardiac specific. Nevertheless, it is useful to begin thinking about the developmental events of heart formation in the context of genetic networks. Here we outline the beginnings of a genetic blueprint for heart development, and consider how perturbations of cardiac developmental control genes may result in different forms of heart disease in humans.

Specification of cardiac cell fate

Studies in model organisms have revealed an evolutionarily conserved programme of heart development, triggered by specific signalling molecules and mediated by tissue-specific transcription factors. This programme controls the genesis of cardiomyocytes from mesodermal stem cells and

the subsequent activation of genes responsible for cardiac contractility and morphogenesis. Cardiomyocytes originate in the anterior lateral mesoderm soon after gastrulation. They are produced in response to protein factors, including bone morphogenetic proteins, that are secreted from adjacent endoderm¹. Cardiogenic signals, which seem to be at least partially conserved across species, activate expression of the homeobox gene *tinman* in flies and the related gene *Nkx2.5* in vertebrates, the earliest molecular markers of the cardiac lineage². *Tinman* is necessary for specification of the cardiac lineage and directly activates transcription of the *Mef2* gene, which encodes a transcription factor that controls myocyte differentiation³. *Tinman* and *Nkx2.5* cooperate with zinc-finger transcription factors of the GATA family to activate cardiac gene expression⁴; these two classes of cardiac transcription factors also regulate each others' expression through mutually reinforcing positive feedback loops⁵.

In contrast to *tinman* in flies, *Nkx2.5* is not essential for specification of the cardiac lineage in mice. Thus, other genes may share related functions with *Nkx2.5*, or cardiogenesis in flies and vertebrates may differ in its dependence on this family of homeobox genes. Functional redundancy between *Nkx2.5* and other cardiac-expressed homeobox genes in vertebrates is suggested by the ability of dominant-negative versions of *Nkx2.5* to block cardiogenesis in frog and zebrafish embryos^{6,7}.

Morphogenesis and looping of the heart tube

Soon after their specification, cardiac muscle cells converge along the ventral midline of the embryo to form a beating

Table 1 Human mutations implicated in congenital heart disease

Disease	Chromosome locus	Gene
Atrial septal defect	5q34	<i>NKX2.5</i>
Atrial septal defect, ventricular septal defect	12q24	<i>TBX5</i>
Pulmonary stenosis; tetralogy of Fallot	20p12	<i>JAGGED-1</i>
Patent ductus arteriosus	6p12	<i>TFAP2B</i>
Supravalvular aortic stenosis	7q11	<i>Elastin</i>
Aortic aneurysm	15q21	<i>Fibrillin</i>
Tetralogy of Fallot; persistent truncus arteriosus; interrupted aortic arch	22q11	?

Box 1

Human genetics in congenital heart disease

We do not yet understand the causes of most complex genetic traits in humans, including congenital heart disease (CHD). For CHD, however, the study of chromosomal disorders and autosomal dominant syndromes, and the genetic linkage analysis of rare pedigrees with milder forms of CHD, have both been informative. A paradigm of variable penetrance and phenotype is emerging, even with single-gene defects, suggesting an important role for secondary factors in CHD.

The 22q11 deletion syndrome provides a way in to understanding the molecular basis of many cardiac neural crest defects. Deletion of a 3-megabase (Mb) region of 22q11 on one copy of chromosome 22 is the most common genetic deletion known in humans and produces DiGeorge or velocardiofacial syndrome⁴². This involves neural crest abnormalities including conotruncal and aortic arch defects, cleft palate and hypoplasia of the thymus and parathyroid glands. It is likely that one or more of the approximately 30 known genes in the 3-Mb 22q11 region is involved in neural crest development. The genetic complexity of this syndrome has, however, precluded definitive identification of the critical gene(s) in this region. Efforts to engineer syntenic chromosome deletions in mice have partially modelled the 22q11 deletion syndrome and are likely to contribute to the identification of the causal gene(s)⁴³. This approach, combined with careful dissection of molecular pathways regulating neural crest-derived cells, should uncover the underlying cause of numerous CHDs⁴⁴.

Recent genetic studies of the cardiac transcription factors NKX2.5 and TBX5 exemplify the synergy between human genetics and studies of model organisms in understanding the aetiology of human CHDs. Many point mutations have been identified in NKX2.5 in families with atrial septal defects and cardiac conduction abnormalities⁴⁵. Sporadic mutations of NKX2.5 have also been found in patients with tetralogy of Fallot, an outflow tract alignment defect, or tricuspid valve anomalies⁴⁶. How these mutations result in cardiac defects is unknown, but the linkage of particular loss-of-function mutations with particular abnormalities suggests that different parts of NKX2.5 function in different developmental decisions in the heart.

Holt–Oram syndrome is characterized by abnormalities in the heart (atrial and ventricular septal defects) and limbs, and is due to mutations in *TBX5* (refs 47, 48). Intriguingly, mutations responsible for

heart and limb defects respectively are clustered in different regions of the protein, suggesting that TBX5 engages different downstream genes or cofactors in the different tissues, and that the distinction again depends on unique structural motifs in the protein⁴⁹. The association of limb and heart defects in Holt–Oram and other syndromes suggests a commonality in developmental control mechanisms for these structures, which are both derived from the lateral mesoderm.

Recent genetic studies of Alagille syndrome revealed a critical role for a well-studied developmental pathway involving the transmembrane receptor, Notch. Alagille syndrome is an autosomal dominant disorder characterized by biliary atresia (absence of biliary ducts) and cardiac defects, typically pulmonary artery stenosis and tetralogy of Fallot. It is caused by mutations in Jagged-1, a ligand for the Notch receptor^{50,51}. Isolated pulmonary stenosis or tetralogy of Fallot have also been associated with Jagged-1 mutations⁵². The Notch signalling pathway is involved in decisions about cell fate and differentiation throughout the embryo, but has only recently been implicated in cardiovascular development. Whether other mediators of Notch signalling, including the gridlock/Hairy-related transcription factor proteins (see main text), are involved in a similar fashion is under investigation.

Char syndrome is an autosomal dominant trait characterized by patent ductus arteriosus (PDA), hand anomalies and facial dysmorphism. Genetic linkage and candidate gene analysis in families with this syndrome revealed dominant-negative mutations in the transcription factor TFAP2B, which is expressed in the neural crest⁴¹. An understanding of ductal regulation through the actions of TFAP2B may lead to insight into the aetiology of isolated cases of PDA, which account for 10% of all CHD.

Analyses of families with supraventricular aortic stenosis (SVAS) and Marfan's syndrome, both autosomal dominant in transmission, have helped understand aortic disease. They are both caused by mutations in structural proteins of connective tissues. SVAS, which is characteristic of William's syndrome, is caused by mutations in elastin, although the pathogenetic mechanism remains unknown⁵³. Mutations in fibrillin result in Marfan's syndrome, which is characterized by progressive ascending aortic aneurysms (dilations)⁵⁴.

linear heart tube composed of distinct myocardial and endocardial layers separated by an extracellular matrix. Studies in mice and fish have revealed an essential role for GATA transcription factors in this process^{8–10}. In all vertebrates, the linear heart tube undergoes rightward looping, which is essential for proper orientation of the pulmonary (right) and systemic (left) ventricles, and for alignment of the heart chambers with the vasculature. The molecular mechanisms governing cardiac looping remain unknown, but identification of genes differentially expressed along the outer and inner curvatures of the looped heart tube suggests that intrinsic properties of the two surfaces may underlie this critical morphogenetic event.

The direction of cardiac looping is determined by an asymmetric axial signalling system that also affects the position of the lungs, liver, spleen and gut (see ref. 11 for a review). Before organ formation begins, this signalling cascade directs the asymmetrical expression of Sonic hedgehog and Nodal, a member of the transforming growth factor- β (TGF- β) family, in the lateral mesoderm. Interpretation of left–right signals is mediated in part by the transcription factor Ptx2, which is expressed along the left side of developing organs, including the early heart tube. Mouse models of left–right defects demonstrate absent, bilaterally symmetrical, or reversed Nodal and Ptx2 expression. In humans, mirror-image reversal of left–right asymmetry is

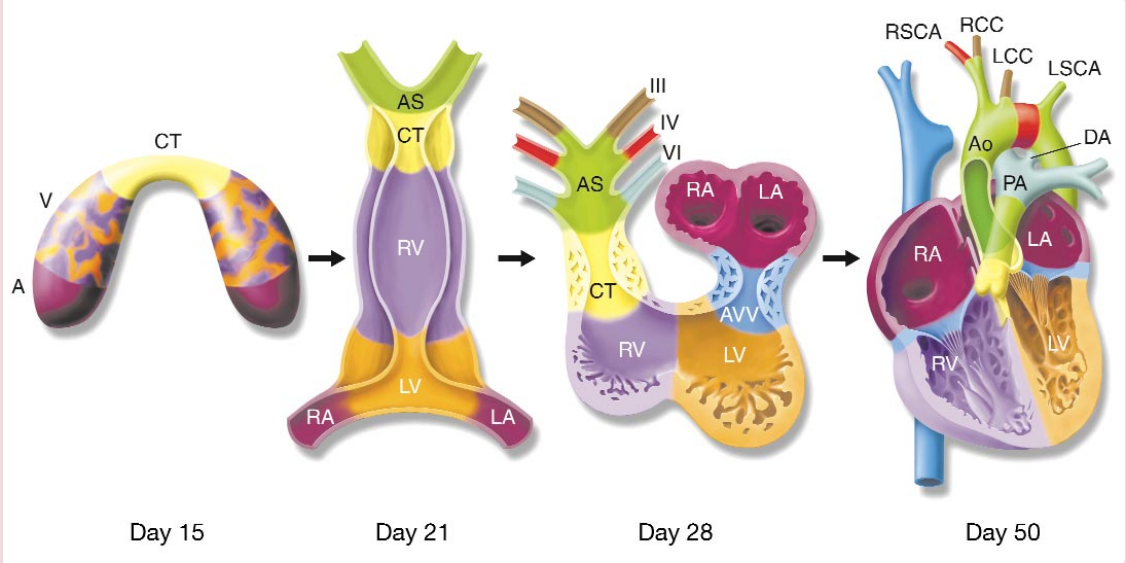
often associated with normal organogenesis. But a discordance of cardiac, pulmonary and visceral asymmetry (heterotaxy syndrome) reflects a lack of coordinated left–right signalling and is universally associated with defects in organogenesis. The common association of human cardiac alignment defects with abnormalities in left–right asymmetry points to intersecting pathways that regulate the direction and process of cardiac looping, and highlights the clinical significance of this area of study.

Segmentation and growth of cardiac chambers

Although individual cardiac chambers do not become morphologically distinguishable until after cardiac looping, their cell fates seem to be genetically programmed much earlier. The linear heart tube is segmentally patterned along the anterior–posterior axis into precursors of the aortic sac, conotruncus (outflow tract), pulmonary and systemic ventricles, and atria (Fig. 1). Each cardiac chamber differs in its morphological and contractile properties and its patterns of gene expression. How chamber identities are established is unknown, but this is likely to involve a combinatorial code of transcription factors, both cardiac-specific and more widely expressed (Fig. 2).

In addition, a subset of ventricular cardiomyocytes surrounding the developing coronary arteries differentiates in response to arterial

Figure 1 Schematic of cardiac morphogenesis. Illustrations depict cardiac development with colour coding of morphologically related regions, seen from a ventral view. Cardiogenic precursors form a crescent (left-most panel) that is specified to form specific segments of the linear heart tube, which is patterned along the anterior–posterior axis to form the various regions and chambers of the looped and mature heart. Each cardiac chamber balloons out from the outer curvature of the



looped heart tube in a segmental fashion. Neural crest cells populate the bilaterally symmetrical aortic arch arteries (III, IV and VI) and aortic sac (AS) that together contribute to specific segments of the mature aortic arch, also colour coded. Mesenchymal cells form the cardiac valves from the conotruncal (CT) and atrioventricular valve (AVV) segments. Corresponding days of human embryonic development are indicated. A, atrium; Ao, aorta; DA, ductus arteriosus; LA, left atrium; LCC, left common carotid; LSCA, left subclavian artery; LV, left ventricle; PA, pulmonary artery; RA, right atrium; RCC, right common carotid; RSCA, right subclavian artery; RV, right ventricle; V, ventricle.

endothelin-1 (ET-1) signalling into tracts of cells that form the cardiac conduction system¹² that regulates heartbeat. The genes regulated by ET-1 that are responsible for this developmental decision are unknown, but are likely to be relevant to cardiac conduction disorders in humans.

Ventricles

Intriguingly, many CHDs affect only a particular segment of the heart, consistent with the notion that each segment develops relatively independently. One of the most severe forms of CHD involves hypoplasia (underdevelopment) of the right or left ventricle. Studies in model organisms have begun to reveal a genetic basis for ventricular development (Fig. 2). The related basic helix–loop–helix (bHLH) transcription factors *dHAND/HAND2* and *eHAND/HAND1* are expressed predominantly in the primitive right and left ventricular segments, respectively, during mouse heart development^{13,14}. Deletion of *dHAND* in mice results in hypoplasia of the right ventricular segment¹⁴. *eHAND* has also been implicated in left ventricular development, although early placental defects of *eHAND* mutant mice precluded a detailed analysis of its role in the heart^{15,16}. Mice lacking *Nkx2.5* also show lethal defects in ventricular morphogenesis and fail to express *eHAND* in the heart, suggesting that *eHAND* may act downstream of *Nkx2.5* to control left ventricular development¹⁷. The ventricular-specific homeobox gene *Irx4* is dependent on *dHAND* and *Nkx2.5* for expression, and is sufficient, when misexpressed in the atria, to activate ventricle-specific gene expression^{18,19}. In the zebrafish, which has a single ventricle, only one *HAND* gene (*dHAND*) has been identified, mutation of which abolishes the ventricular segment of the heart²⁰. Mice lacking the transcription factor *MEF2C*, normally expressed throughout the atrial and ventricular chambers, also show hypoplasia of the right and left ventricles, resulting in their early demise in embryogenesis²¹. The cardiac defects in *MEF2C* mutant embryos suggest that *MEF2C* might be a necessary cofactor for one or more ventricular-restricted regulatory proteins. Whether members of such ventricular developmental pathways are disrupted in human syndromes of right or left ventricular hypoplasia remains to be determined.

Atria

Considerably less is known about the genetics of atrial development. The orphan nuclear receptor COUP-TFII is expressed in atrial

precursors and is required for atrial, but not ventricular, growth²². Retinoid signalling has also been implicated in atrial specification and in the regulation of the atrioventricular border along the anterior–posterior axis of the heart tube²³. How this border is established remains unclear. In Ebstein's anomaly in humans, the right atrioventricular valve is displaced inferiorly into the ventricle, resulting in 'atrialization' of ventricular myocardium. It will be interesting to determine whether the gene responsible for Ebstein's anomaly is involved in atrioventricular specification.

Growth

Proliferation of cardiomyocytes within each chamber is necessary to support the increasing haemodynamic load during embryonic development. Neuregulin growth factors, secreted from the endocardium, and their myocardial receptors ErbB2 and ErbB4, are required for development of trabeculae, the finger-like projections comprised of the ventricular cardiomyocytes²⁴. Defects in ventricular trabeculation have also been observed in mice lacking angiogenic factors, such as vascular endothelial growth factor (VEGF)²⁵ and angiopoietin-1 (ref. 26), that are expressed in the endocardium. Understanding the mechanism of reciprocal signalling between endocardium and myocardium may provide clues to the regulation of myocardial cell proliferation during development. Harnessing this information to drive adult myocardial cells into a proliferative state remains one of the most important and challenging goals in this field.

Cardiac valve formation

Appropriate placement and function of cardiac valves is essential for division of the chambers and to ensure the unidirectional flow of blood through the heart. Early in development, septation of the cardiac tube into distinct chambers is achieved through regional swellings of extracellular matrix, known as cardiac cushions, that form the anlage of atrioventricular and ventriculoarterial valves (Fig. 1). Reciprocal signalling between the endocardial and myocardial cell layers in the cushion region, mediated in part by TGF- β family members, induces a transformation of endocardial cells into mesenchymal cells²⁷. These migrate into the cushions and differentiate into the fibrous tissue of the valves; they are also involved in septation of the common atrioventricular canal into right- and left-sided

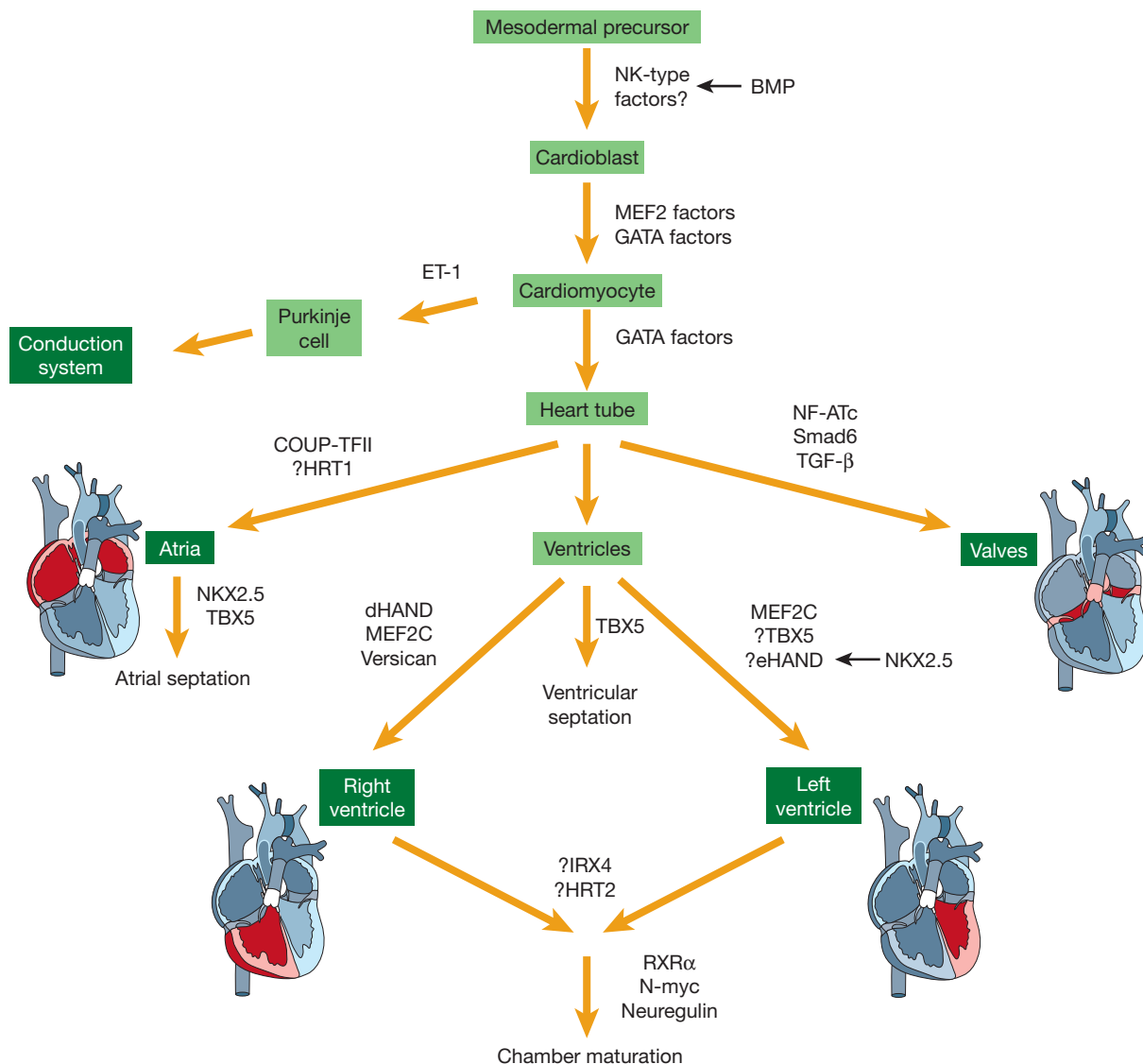


Figure 2 A genetic blueprint for heart development. This schematic shows particular steps in cardiac morphogenesis, focusing on mesodermal contributions. Genetic pathways leading to linear heart-tube formation are partly conserved between *Drosophila* and mouse. The formation of valves, ventricles, atria and the conduction system are under the control of groups of regulatory proteins that may act independently or in a common pathway. Cardiac regions affected by specific pathways are highlighted in dark green boxes with white type. Factors necessary for distinct steps during cardiogenesis in model organisms or humans are indicated beside the arrows. Regulatory factors suspected to have region-specific roles are indicated with a '?'. Cell types or regions of the heart are indicated in boxes. Distinct processes during cardiogenesis (atrial septation, ventricular septation and chamber maturation) are indicated.

orifices. Trisomy 21 (Down syndrome) in humans is commonly associated with incomplete septation of the atrioventricular valves, but the gene(s) on chromosome 21 responsible for valve development remains unknown. Abnormalities in extracellular matrix composition, production or reabsorption may underlie many such defects.

Gene targeting in mice has revealed important roles for the NF-ATc and Smad6 transcription factors in the formation of cardiac valves. Mice lacking NF-ATc, a downstream mediator of signalling by the calcium-dependent protein phosphatase calcineurin, exhibit fatal defects in valve formation, reflecting a potential role for calcineurin in transduction of signals for valvulogenesis^{28,29}. Smad6, which is implicated in the activation of gene expression in response to TGF-β signalling, is also expressed specifically in cardiac valve precursors, but disruption of *Smad6* leads to abnormally thickened, gelatinous valves³⁰. Interestingly, most human valve defects, which are among the most common CHDs, result in stenotic

(narrow) or incompetent valves that are often found to be thick and gelatinous on autopsy.

Vascular connections to the heart

During the evolutionary transition from aquatic to terrestrial organisms, separation of the pulmonary and systemic circulations into two parallel circuits became necessary. This was accomplished in part by cardiac neural crest cells that migrate from the neural folds into the pharyngeal arches and heart, where they have essential roles in septation of the single outflow of the heart (the truncus arteriosus) into the aorta and pulmonary artery, and in the formation of the conotruncal portion of the ventricular septum³¹ (Fig. 1). After septation of the aorta and pulmonary artery, the vessels rotate in a twisting fashion to achieve their final connections with the left and right ventricles, respectively. Neural crest cells also contribute to the bilaterally symmetrical aortic arch arteries that arise from the aortic sac and undergo extensive remodelling, resulting in formation of the aorta,

ductus arteriosus, proximal subclavian, and carotid and pulmonary arteries (Fig. 1). Perturbation of neural crest cell specification or migration can result in life-threatening abnormalities of this region, which account for about 30% of congenital heart abnormalities in humans.

Mice lacking ET-1 or its G-protein-coupled receptor, ETA, have post-migratory cardiac neural crest defects, cleft palate and other craniofacial anomalies reminiscent of 22q11 deletion syndrome in humans^{32,33} (see Box 1). *dHAND* and *eHAND*, normally expressed in the neural crest-derived pharyngeal and aortic arches, are downregulated in these structures in mice deficient in *ET-1* or *ETA*, suggesting that the HAND proteins are regulated by ET-1 signalling³⁴. Consistent with this notion, *dHAND*-null mice also exhibit a severe defect in survival of pharyngeal and aortic arch mesenchyme³⁴. *Neuropilin-1*, a semaphorin and VEGF receptor, is downregulated in *dHAND* mutants, and targeted mutation of *neuropilin-1* results in a phenotype similar to that of *ET-1* mutants. This suggests that ET-1, *dHAND* and *neuropilin-1* may function in a common pathway regulating neural crest development^{35,36}.

Much like the heart itself, segments of the mature aortic arch are derived from distinct embryonic structures (aortic arches III, IV and VI, see Fig. 1) that develop relatively independently. Disruption of the Forkhead transcription factor *Mfh1* causes hypoplasia of the fourth aortic arch artery in mice, resulting in absence of the transverse aortic arch³⁷. This phenotype resembles the interruption of the aortic arch seen in humans. Mice lacking combinations of retinoic acid receptors (RAR/RXR) or the homeobox gene *pax3* also have a variety of outflow tract and aortic arch defects^{38,39}. In zebrafish, mutation of the *gridlock* gene, which is expressed in the arch arteries and encodes a bHLH transcription factor related to the *Drosophila* protein *Hairy*, results in coarctation or narrowing of a particular region of the aorta, a defect commonly observed in humans⁴⁰. This may reflect an essential role for *gridlock* in controlling the decision of an early vascular progenitor to adopt an arterial fate in a specific vascular segment. Finally, evidence for independent regulation of the sixth aortic arch artery comes from the third most common CHD — patent ductus arteriosus — in which the ductus arteriosus, an embryonic vessel connecting the aorta and pulmonary artery, fails to close after birth. Heterozygous mutations of the transcription factor *TFAP2B* can result in familial patent ductus arteriosus, suggesting a role for *TFAP2B* in governing closure of this vessel⁴¹.

Looking ahead

Given the complexity of cardiac development and the devastating consequences of even subtle perturbations in the process, it is not surprising that mutations in such a large number of genes can cause cardiac malformations (Table 2). The heterogeneity of CHDs associated with single-gene defects, as demonstrated for *NKX2.5* or *TBX5* mutations, makes mechanistic understanding of gene function challenging and points to the importance of modifier genes, environmental factors and genetic polymorphisms in determining the severity and type of CHD. Clinically, the prospect that secondary factors affect the ultimate phenotype of heterozygous genetic mutations provides hope that at least some types of heart defects might be prevented by modulating these cofactors.

Elucidation of the genetic networks and mechanisms underlying cardiac development also opens up the possibility of overcoming the inability of adult cardiomyocytes to divide, which is one of the major obstacles in cardiovascular medicine in adults. Genetic strategies for converting non-muscle cells to a cardiac cell fate through expression of developmental control genes that specify cardiac cell identity might eventually be used in cardiac regeneration or cell replacement.

Finally, the availability of complete genome sequences for humans and model organisms should revolutionize our understanding of cardiac development. Mapping of human CHD genes, once daunting, will be simplified, and genomic sequence profiling will enable genetic screening for mutations and polymorphisms in CHD

Table 2 Cardiac defects associated with gene mutations

Protein	Species	Cardiac defect
Signalling molecule		
Neuregulin	Mouse	Absent myocardial trabeculation
ErbB2, B4	Mouse	Absent myocardial trabeculation
Neurotrophin-3	Mouse	Outflow tract defects
Neuropilin-1	Mouse	Outflow tract and arch defects
Endothelin-1	Mouse	Outflow tract and arch defects
Jagged-1	Human	Pulmonary stenosis, outflow tract
TGF- β	Mouse	Cardiac valve defects
Endoglin	Human	Arteriovenous malformations
Cell adhesion molecule		
VCAM	Mouse	Epicardial dissolution
$\alpha 4$ integrin	Mouse	Epicardial dissolution
Versican	Mouse	Hypoplastic right ventricle
Ion channel		
HERG	Human	Long QT, Brugada syndrome
SCN5A	Human	Long QT syndrome
KVLQT1	Human	Long QT syndrome
Mink	Human	Long QT syndrome
Transcription factor		
GATA-4	Mouse	Cardiac bifida
GATA-5	Zebrafish	Cardiac bifida
Fog-2	Mouse	Cardiac alignment, coronary artery defects
MEF2C	Mouse	Hypoplastic right and left ventricles
dHAND	Mouse, zebrafish	Hypoplastic right ventricle, arch defects
COUP-TFII	Mouse	Hypoplastic atria
Nkx2.5	Mouse, Human	Looping, conduction defect, atrial septal defect
TBX5	Human	Ventricular septal defect, atrial septal defect
NF-ATc	Mouse	Absent valves
Smad6	Mouse	Thickened valves
Pax3	Mouse	Persistent truncus arteriosus
RXR/RAR	Mouse	Outflow tract, aortic arch, myocardial defects
Mfh1	Mouse	Interrupted aortic arch
Gridlock/HRT2	Zebrafish	Coarctation of aorta
NF-1	Mouse	Outflow tract defects, thin myocardium
TFAP2B	Human	Patent ductus arteriosus
N-Myc	Mouse	Thin myocardium
TEF-1	Mouse	Thin myocardium
WT-1	Mouse	Epicardial dissolution

genes. Comparative analysis of genes involved in cardiac development and disease in different species will enable human disease genes to be identified from their orthologues in other species, and vice versa. The identification of genes acting downstream in cardiac developmental hierarchies, which is a major challenge at present, will be made easier by the ability to profile gene expression in the normal and abnormal heart. As we enter the post-genomic era, the emerging genetic blueprint of heart development will enable us to understand the all too common complex genetic trait that is CHD. □

- Schultheis, T. M., Xydas, S. & Lassar, A. B. Induction of avian cardiac myogenesis by anterior endoderm. *Development* **121**, 4203–4214 (1995).
- Harvey, R. P. NK-2 homeobox genes and heart development. *Dev. Biol.* **178**, 203–216 (1996).
- Gajewski, K., Kim, Y., Lee, Y. M., Olson, E. N. & Schulz, R. A. *D-Mef2*: a target for tinman activation during *Drosophila* heart development. *EMBO J.* **16**, 515–522 (1998).
- Durocher, D., Charron, F., Warren, R., Schwartz, R. J. & Nemer, M. The cardiac transcription factors *Nkx2-5* and *GATA-4* are mutual cofactors. *EMBO J.* **16**, 5687–5696 (1997).
- Schwartz, R. J. & Olson, E. N. Building the heart piece by piece: modularity of cis elements regulating *Nkx2.5* transcription. *Development* **126**, 4187–4192 (1999).
- Fu, Y., Yan, W., Mohun, T. J. & Evans, S. M. Vertebrate tinman homologues *XNkx2-3* and *XNkx2-5* are required for heart formation in a functionally redundant manner. *Development* **125**, 4439–4449 (1998).
- Grow, M. W. & Kreig, P. A. Tinman function is essential for vertebrate heart development: elimination of cardiac differentiation by dominant inhibitory mutants of the tinman-related genes,

- XNkx2-3 and XNkx2-5. *Dev. Biol.* **204**, 187–196 (1998).
8. Molkentin, J., Lin, Q., Duncan, S. A. & Olson, E. N. Requirement of the GATA4 transcription factor for heart tube formation and ventral morphogenesis. *Genes Dev.* **11**, 1061–1072 (1997).
9. Kuo, C. T. *et al.* GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev.* **11**, 1048–1060 (1997).
10. Reiter, J. F. *et al.* *Gata5* is required for the development of the heart and endoderm in zebrafish. *Genes Dev.* **13**, 2983–2995 (1999).
11. Capdevila, J., Vogan, K. J., Tabin, C. J. & Belmonte, J. C. Mechanisms of left–right determination in vertebrates. *Cell* **101**, 9–21 (2000).
12. Hyer, J. *et al.* Induction of Purkinje fiber differentiation by coronary arterialization. *Proc. Natl Acad. Sci. USA* **96**, 13214–13218 (1999).
13. Srivastava, D., Cserjesi, P. & Olson, E. N. New subclass of bHLH proteins required for cardiac morphogenesis. *Science* **270**, 1995–1999 (1995).
14. Srivastava, D. *et al.* Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nature Genet.* **16**, 154–160 (1997).
15. Firulli, A. B., McFadden, D. G., Lin, Q., Srivastava, D. & Olson, E. N. Heart and extra embryonic mesodermal defects in mouse embryos lacking the bHLH transcription factor Hand1. *Nature Genet.* **8**, 266–270 (1998).
16. Riley, P., Anson-Cartwright, L. & Cross, J. C. The Hand1 bHLH transcription factor is essential for placental and cardiac morphogenesis. *Nature Genet.* **18**, 271–275 (1998).
17. Biben, C. & Harvey, R. P. Homeodomain factor Nkx2-5 controls left/right asymmetric expression of bHLH gene *eHand* during murine heart development. *Genes Dev.* **11**, 1357–1369 (1997).
18. Bao, Z., Bruneau, B. G., Seidman, J. G., Seidman, C. E. & Cepko, C. L. Regulation of chamber-specific gene expression in the developing heart by *Irx4*. *Science* **283**, 1161–1164 (1999).
19. Bruneau, B. G. *et al.* Cardiac expression of the ventricle-specific homeobox gene *Irx4* is modulated by Nkx2-5 and dHand. *Dev. Biol.* **217**, 266–277 (2000).
20. Yelon, D. *et al.* Parallel roles for the bHLH transcription factor HAND2 in zebrafish and pectoral fin development. *Development* **127**, 2573–2582 (2000).
21. Lin, Q., Schwarz, J., Bucana, C. & Olson, E. N. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* **276**, 1404–1407 (1997).
22. Pereira, F. A., Qui, Y., Zhou, G., Tsai, M. & Tsai, S. The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development. *Genes Dev.* **13**, 1037–1049 (1999).
23. Dyson, E. *et al.* Atrial-like phenotype is associated with embryonic ventricular failure in retinoid X receptor alpha-/- mice. *Proc. Natl Acad. Sci. USA* **92**, 7386–7390 (1995).
24. Lee, K. F. *et al.* Requirement for neuregulin receptor erbB2 in neural and cardiac development. *Nature* **378**, 394–398 (1995).
25. Carmeliet, P. *et al.* Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* **380**, 435–439 (1996).
26. Suri, C. *et al.* Requisite role of angiopoietin-1, a ligand for the TIE2 receptor, during embryonic angiogenesis. *Cell* **87**, 1171–1180 (1996).
27. Brown, C. B., Boyer, A. S., Runyan, R. B. & Barnett, J. V. Requirement of type III TGF beta receptor for endocardial cell transformation in the heart. *Science* **283**, 2080–2082 (1999).
28. Ranger, A. M. *et al.* The transcription factor NF-ATc is essential for cardiac valve formation. *Nature* **392**, 186–190 (1998).
29. De la Pompa, J. L. *et al.* Role of the NF-ATc transcription factor in morphogenesis of cardiac valves and septum. *Nature* **392**, 182–186 (1998).
30. Galvin, K. M. *et al.* A role for smad6 in development and homeostasis of the cardiovascular system. *Nature Genet.* **24**, 171–174 (2000).
31. Kirby, M. L. & Waldo, K. L. Neural crest and cardiovascular patterning. *Circ. Res.* **77**, 211–215 (1995).
32. Kurihara, Y. *et al.* Aortic arch malformations and ventricular septal defect in mice deficient in endothelin-1. *J. Clin. Invest.* **96**, 293–300 (1995).
33. Clouthier, D. E. *et al.* Cranial and cardiac neural crest defects in endothelin-A receptor-deficient mice. *Development* **125**, 813–824 (1998).
34. Thomas, T. *et al.* A signaling cascade involving endothelin-1, dHAND and Msx1 regulates development of neural crest-derived branchial arch mesenchyme. *Development* **125**, 3005–3014 (1998).
35. Yamagishi, H., Olson, E. N. & Srivastava, D. The bHLH transcription factor, dHAND, is required for vascular development. *J. Clin. Invest.* **105**, 261–270 (2000).
36. Kawasaki, T. *et al.* A requirement for neuropilin-1 in embryonic vessel formation. *Development* **126**, 4854–4902 (1999).
37. Iida, K. *et al.* Essential roles of the winged helix transcription factor MFH-1 in aortic arch patterning and skeletogenesis. *Development* **124**, 4627–4638 (1997).
38. Li, J., Liu, K. C., Jin, F., Lu, M. M. & Epstein, J. A. Transgenic rescue of congenital heart disease and spina bifida in Splotch mice. *Development* **126**, 2495–2503 (1999).
39. Gruber, P. J. *et al.* RXR alpha deficiency confers genetic susceptibility for aortic sac, conotruncal, atrioventricular cushion, and ventricular muscle defects in mice. *J. Clin. Inv.* **98**, 1332–1343 (1996).
40. Zhong, T. P., Rosenberg, M., Mohideen, M. A., Weinstein, B. & Fishman, M. C. Gridlock, an HLH gene required for assembly of the aorta in zebrafish. *Science* **287**, 1820–1824 (2000).
41. Satoda, M. *et al.* Mutations in TFAP2B cause Char syndrome, a familial form of patent ductus arteriosus. *Nature Genet.* **25**, 42–46 (2000).
42. Emanuel, B. S., Budarf, M. L., & Scambler, P. J. in *Heart Development* (eds Harvey, R. P. & Rosenthal, N.) 463–478 (Academic, New York, 1998).
43. Lindsay, E. A. *et al.* Congenital heart disease in mice deficient for the DiGeorge syndrome region. *Nature* **401**, 379–383 (1999).
44. Yamagishi, H., Garg, V., Matsuoka, R., Thomas, T. & Srivastava, D. A molecular pathway revealing a genetic basis for human cardiac and craniofacial defects. *Science* **283**, 1158–1161 (1999).
45. Schott, J.-J. *et al.* Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* **281**, 108–111 (1998).
46. Benson, D. W. *et al.* Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J. Clin. Invest.* **104**, 1567–1573 (1999).
47. Basson, C. T. *et al.* Mutations in human TBX5 cause limb and cardiac malformation in Holt-Oram syndrome. *Nature Genet.* **15**, 30–35 (1997).
48. Li, Q. Y. *et al.* Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nature Genet.* **15**, 21–29 (1997).
49. Basson, C. T. *et al.* Different TBX5 interactions in heart and limb defined by Holt-Oram syndrome mutations. *Proc. Natl Acad. Sci. USA* **96**, 2919–2924 (1999).
50. Li, L. *et al.* Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nature Genet.* **16**, 243–251 (1997).
51. Oda, T., Elkhoul, A. G. & Pike, B. L. Mutations in the human Jagged1 gene are responsible for Alagille syndrome. *Nature Genet.* **16**, 235–242 (1997).
52. Krantz, I. D. *et al.* Jagged1 mutations in patients ascertained with isolated congenital heart defects. *Am. J. Hum. Genet.* **84**, 56–60 (1999).
53. Curran, M. E. *et al.* The elastin gene is disrupted by a translocation associated with supravalvular aortic stenosis. *Cell* **73**, 159–168 (1993).
54. Dietz, H. C. *et al.* Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* **352**, 279–281 (1991).

Acknowledgements

D.S. is supported by grants from NIH, American Heart Association, March of Dimes and Smile Train. E.N.O. is supported by grants from NIH, Muscular Dystrophy Association, The Robert A. Welch Foundation and the D. W. Reynolds Foundation. The authors thank R. Sahadi and S. Johnson for assistance with graphics.

Genomic circuits and the integrative biology of cardiac diseases

Kenneth R. Chien

UCSD-Salk Program in Molecular Medicine and the UCSD Institute of Molecular Medicine, La Jolla, California 92093, USA
(e-mail: kchien@ucsd.edu)

Human cardiac disease is the result of complex interactions between genetic susceptibility and environmental stress. The challenge is to identify modifiers of disease, and to design new therapeutic strategies to interrupt the underlying disease pathways. The availability of genomic databases for many species is uncovering networks of conserved cardiac-specific genes within given physiological pathways. A new classification of human cardiac diseases can be envisaged based on the disruption of integrated genomic circuits that control heart morphogenesis, myocyte survival, biomechanical stress responses, cardiac contractility and electrical conduction.

Our understanding of the molecular circuits that drive the onset of complex cardiac diseases remains primitive. The paucity of informative human mutations, the high frequency of embryonic lethality in gene-targeted mouse models, and the complexity of physiological endpoints have made it difficult to uncover new disease pathways. Because of a lack of reliable surrogate models of disease progression, new therapeutic strategies must rely on the crude endpoint of survival in mega-scale clinical trials, and cardiovascular drug discovery is now viewed as an extremely high-risk enterprise. Clearly, the field of cardiac biology is in transition: new technology must be harnessed to explore uncharted waters—to adopt an ‘offshore’ approach to medical discovery. Fortunately, recent advances in cardiovascular genomics are leading to a resurgence in the integrative biology of complex cardiac diseases.

Morphogenetic circuits and congenital heart defects

The genetics of congenital heart disease (CHD) point to the existence of powerful disease modifiers. A wide phenotypic spectrum is seen in patients harbouring identical disease alleles¹ and in mutant mice bred into varying genetic backgrounds^{2,3}. The variation in phenotypic penetrance and severity suggests that if we can identify high-risk individuals, a reduction in infant morbidity might be possible by altering environmental or maternal factors. This risk-reduction strategy has already led to a marked reduction in neural tube defects by the augmentation of maternal dietary folate⁴. Although cardiac muscle defects accompany many types of CHD, they often arise as secondary effects of a primary disturbance in non-muscle cell lineages, such as endothelium and neural crest^{5,6}. Understanding the paracrine signalling pathways of specialized non-myocytes may be crucial in reducing the risk of CHD.

The formation of the outflow tracts and valves of the heart requires the migration of neural crest cells into the embryonic heart (see review in this issue by Srivastava & Olson, pages 221–226), and genetic analysis of cardiac neural crest defects is uncovering new paradigms for CHD (Fig. 1, Table 1). Neural crest defects account for over 10% of all CHD, and are often associated with extra-cardiac anomalies⁷. The

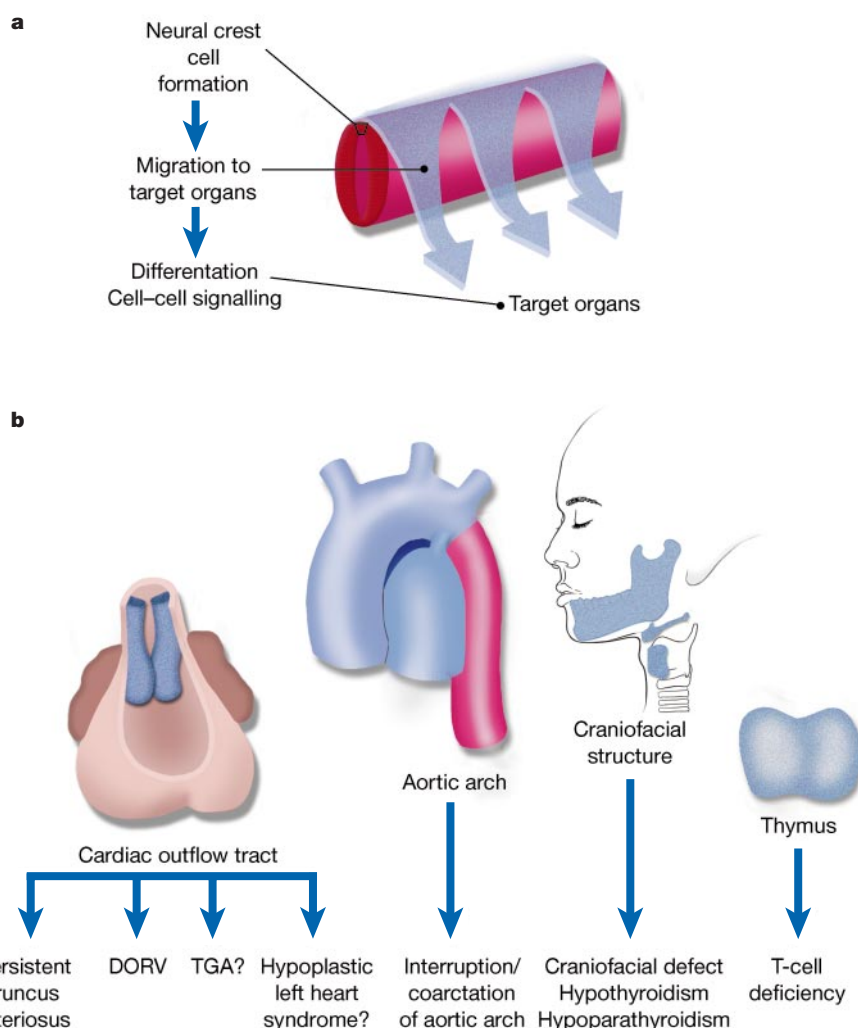
spectrum of defects may reflect particular deficiencies in neural crest formation, migration, transition to smooth muscle, and/or paracrine signalling to cardiac and cushion mesenchymal cell lineages. Studies of DiGeorge syndrome patients harbouring microdeletions of chromosome 22 have identified a minimal genetic interval for this disease, encompassing 300 kilobases and more than 10 genes⁸. Although a clinical study of one patient implicated a single gene involved in ubiquitination (*UFD1L*)⁹, studies in Cre-lox engineered mice now point to a different locus¹⁰. Mice with a 1.2-megabase deletion in the region of chromosome 16 syntenic with human chromosome 22 display neural crest defects identical to those in a subset of DiGeorge patients¹⁰. Genetic complementation has narrowed the disease interval to two genes (A. Baldini, personal communication). Because DiGeorge syndrome usually includes defects in non-cardiac tissues (thymus and craniofacial abnormalities), the genetic lesion may lie in an early step of neural crest formation or migration (Fig. 1). Endothelin-1 (ref. 11), vitamin A signals^{2,3} and other pathways have been linked to neural crest defects (Table 1). Gene targeting restricted to cardiovascular cell-specific lineages³ will be valuable in dissecting the particular roles of these widely expressed genes in forms of CHD.

The effects of ablating neural crest in the chick embryo suggest that neural crest cells send paracrine signals that regulate the growth and function of embryonic cardiomyocytes¹². New paracrine pathways should be elucidated by ‘signal-sequence trapping’ to identify secreted factors and the generation of databases of the genes expressed in neural crest lineages. Validating the functions of these genes in mouse models (Table 1) should identify new candidate genes for hypoplastic left heart syndromes (HLHS)⁷—a life-threatening CHD that includes severe defects both in the outflow tract and in ventricular growth. Interestingly, the association of aortic valve atresia (the complete absence of the aortic valve) with HLHS suggests a possible neural crest origin. Higher-throughput candidate gene analyses in sporadic cases of hypoplastic heart disease may uncover new genes for this devastating form of CHD.

Survival circuits and initiation of heart failure

Heart failure is a leading cause of mortality worldwide¹³. Patients receive symptomatic treatment, and biologically

Figure 1 Cardiac neural crest defects. **a**, Biological pathways for cardiac neural crest-related defects. **b**, Neural crest-related disease phenotypes. DORV, double outlet right ventricle; TGA, transposition of the great arteries.



targeted therapy will rest on the discovery of new pathways that initiate, promote or potentially reverse the onset of heart-muscle failure in response to stress¹⁴.

The failure of cell-survival pathways to inhibit myocyte apoptosis is a critical step in the initiation of heart failure (Fig. 2)¹⁴. Cardiotrophin-1, a gp130 receptor-dependent cytokine, was isolated from a mouse embryonic stem-cell model of cardiogenesis, and is a potent cardiac myocyte survival factor¹⁵. Mice with a cardiac-restricted ablation of gp130 function cannot activate downstream survival pathways and display massive apoptosis of myocytes, acute heart failure and high mortality in response to moderate hypertension¹⁶. Cardiotrophin-1 levels are raised in human heart failure, and studies in the intact heart indicate a major cardioprotective effect for cardiotrophin-1 during hypoxic injury, supporting its potential clinical relevance. Another paracrine factor, neuregulin, can activate survival pathways in both ventricular and neuronal cells *in vitro*¹⁷. Ablation of neuregulin or its receptors, ErbB2 and ErbB4, leads to heart failure in the mouse embryo¹⁸. Clinical studies in patients receiving anti-ErbB2 antibodies to treat metastatic breast cancer have uncovered evidence of cardiomyopathy in a small subset of patients¹⁹. Ongoing studies of mice with cardiac-restricted ablation of ErbB2 function should be critical in determining whether neuregulin also has a role in survival pathways in the adult heart *in vivo*. As neuregulin is an endothelial peptide¹⁸, a search for new endothelial-derived myocyte survival factors may be warranted. Endothelial expression databases should reveal new candidate

myocyte survival factors that are worth evaluating in models of cardiomyopathy.

Biomechanical circuits and dilated cardiomyopathy

Dilation of the ventricular chamber and the associated increase in stress on the ventricular wall are often the first irreversible steps towards heart failure. Accordingly, the mechanisms that link biomechanical forces to the activation of stress pathways are central to understanding the progression of heart failure²⁰. In isolated myocytes, mechanical stretch can directly trigger growth signals, suggesting a role for a biomechanical sensor. Although cardiomyocytes have classical stretch receptors, these may not mediate stretch-activated responses²⁰. Studies in human and mouse models of dilated cardiomyopathy, in which the heart chamber markedly enlarges, now point to a critical role for muscle-cell Z-disc components in chamber dilation²¹ (Fig. 2), suggesting that the sensor might be located at the interface of the cytoskeleton and sarcomere.

An initial clue to the importance of Z-disc components in dilated cardiomyopathy came from mice lacking the muscle LIM-domain protein (MLP), a Z-disc protein that interacts directly with α -actinin. MLP-deficient mice display many features of human dilated cardiomyopathy²². Mutations in other proteins associated with α -actinin — cardiac α -actin, desmin and titin — were subsequently found in familial forms of human dilated cardiomyopathy^{23–25}. In the mouse, mutations in other Z-disc components result in dilated cardiomyopathy^{14,21}. Recently, mice mutant for α -actinin-associated

Table 1 Single-gene mutations that cause neural crest-related phenotypes

Gene mutated	Expression	Aortic arch defects	PTA	TGA	DORV or TOF	Thymus defects	Craniofacial defects	Other
<i>Pax3</i> (ref. 40)	Neural crest; neural tube	+	+	–	+	+	+	Lethal before embryonic stage (e)13
<i>ETA</i> (ref. 41) or <i>ET-1</i> (ref. 42) or <i>ECE-1</i> (ref. 43)	ETA: neural crest; ET-1: endothelium; ECE-1: mesenchyme, endocardial cells	+	+	+	+	+	+	
<i>Sox4</i> (ref. 44)	Endocardium at outflow, atrioventricular cushion	Rare	+	+	+	–	–	Lethal at e14; hypoplastic left ventricle
<i>RXRα</i> (ref. 2)	Neural crest, myocardium, endocardium, epicardium	–	+	–	+	–	–	Lethal at e14; thin myocardium, eye defect
<i>Cx43</i> (ref. 45)	Neural crest, myocardium	–	–	–	–	+	–	Pulmonary stenosis
<i>NT-3</i> (ref. 46) or <i>TrkC</i> (ref. 47)	Neural crest, myocardium, smooth muscle	–	+	–	+	–	–	Pulmonary stenosis, ASD, VSD
<i>NF-1</i> (ref. 48)	Neural crest	–	–	–	+	–	–	Hyperplasia in outflow tract, thin myocardium
<i>NMHC-B</i> (ref. 49)	Neural crest, myocardium	–	–	–	+	–	–	Pulmonary stenosis, ventricular hypertrophy
<i>TGF-β2</i> (ref. 50)	Myocyte, cushion mesenchyme	Aortic wall thinning	–	–	+	–	+	
<i>dHAND</i> (ref. 51)	Neural crest, myocardium	+	–	–	–	–	+	Lethal at e11; loss of right ventricle
<i>MFH-1</i> (ref. 52)	Neural crest, mesenchyme	+	–	–	–	–	+	VSD
<i>Rae28</i> (ref. 53)	Neural crest?	–	–	–	+	+	+	
DiGeorge syndrome (ref. 54)	Gene not yet determined (chromosome 22q11)	+	+	Rare	+	+	+	Pulmonary stenosis, VSD

Abbreviations: ASD, atrial septal defect; DORV, double outlet right ventricle; PTA, persistent truncus arteriosus; TGA, transposition of the great arteries; TOF, tetralogy of Fallot; VSD, ventricular septal defect.

LIM-domain protein have pointed to a critical role for this protein in the onset of a form of right ventricular cardiomyopathy (M. Pashforoush and K.R.C., unpublished results). Defects in the human plakoglobin gene can cause a recessive form of right ventricular dysplasia, again pointing to a potential importance of α -actinin complexes at the Z-line²⁶. A model can now be proposed in which biomechanical stress leads to defects in Z-disc formation, sarcomeric stabilization and muscle-cell signalling, thereby driving the onset of

chamber dilation and heart failure. Genomic databases contain a number of novel muscle-restricted cytoskeletal proteins which might be components of the macromolecular Z-disc complex. The crystallization and structural determination of α -actinin domains should be valuable in defining how associated proteins influence α -actinin function in living cardiac cells²⁷.

The convergence of findings in humans and mouse models might suggest a single unifying molecular pathway. But there are a diversity

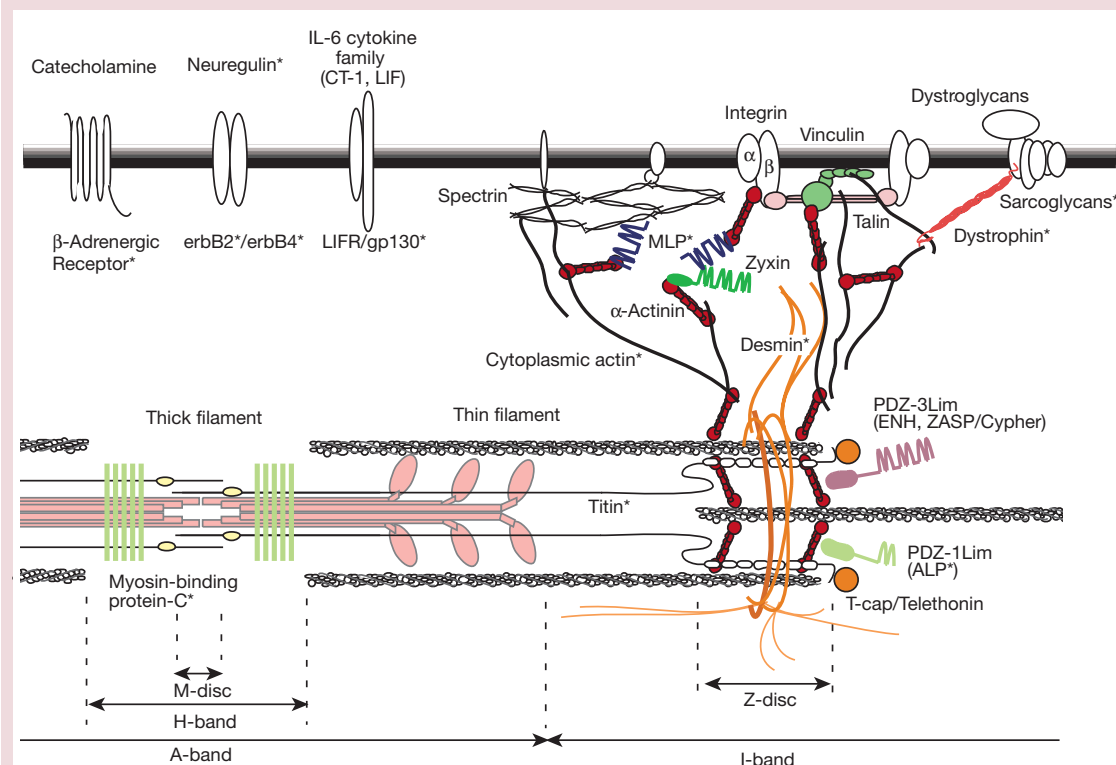
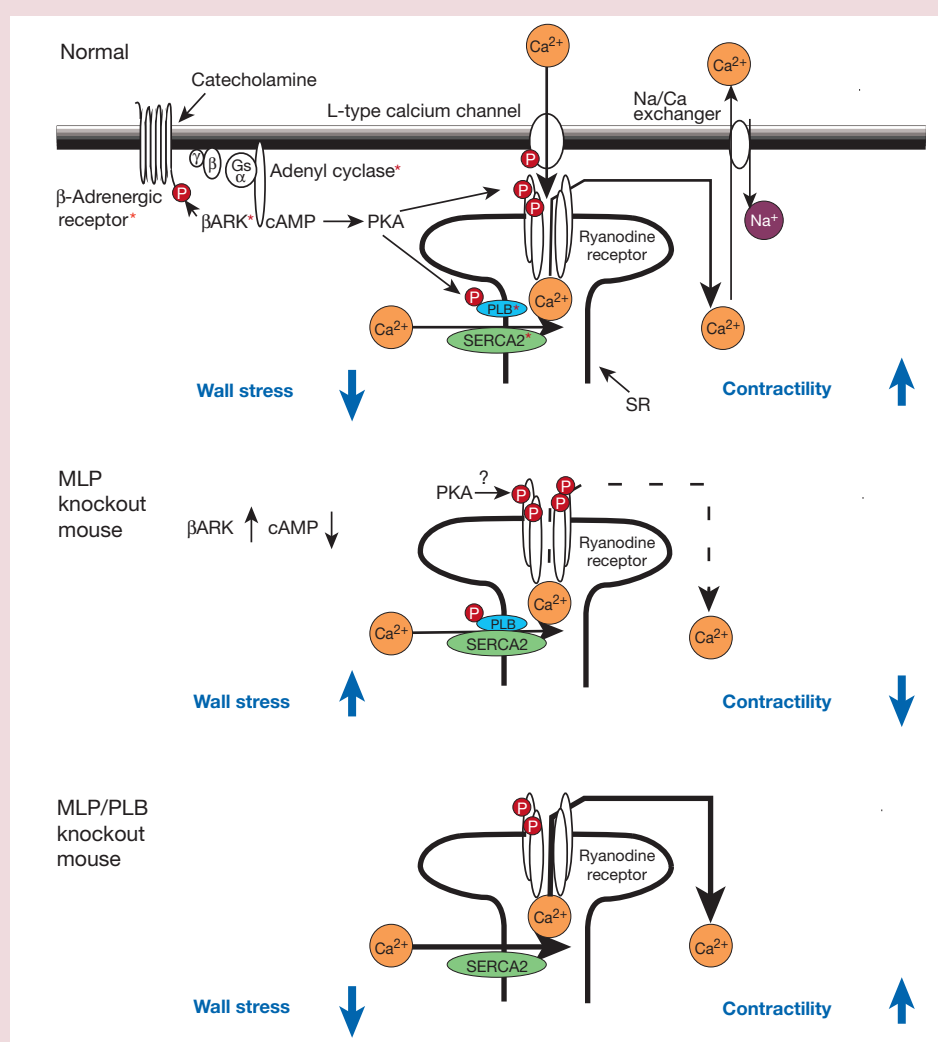


Figure 2 Genetic modifiers and pathways for dilated cardiomyopathy. Asterisks indicate genes that are either human disease genes and/or strong genetic *in vivo* modifiers of dilated cardiomyopathy and/or heart failure in gene-targeted mice.

Figure 3 Calcium cycling pathways in a cardiomyocyte and their involvement in heart failure. **a**, In the normal heart, β -adrenergic-receptor stimulation results in an increase in contractility and relaxation by leading to phosphorylation of the muscle-specific inhibitor (phospholamban) of the sarcoplasmic reticulum (SR) calcium ATPase pump (SERCA), which enhances pump activity. The consequent release of Ca^{2+} from stores in the SR leads to an increase in muscle-cell contractility. Ca^{2+} is removed from the cytosol by re-entering the SR through SERCA2, leading to relaxation. **b**, In the cardiomyopathic heart of MLP $^{-/-}$ mice, the β -adrenergic pathway is blunted, resulting in decreased phospholamban phosphorylation and decreased SERCA activity. The decrease in calcium cycling leads to decreased contractility and relaxation. **c**, In mice having a double mutation in both MLP and PLB, the SERCA activity is released from the inhibition by PLB. The β -adrenergic pathway is short-circuited and contractility and relaxation are increased independent of increases in cAMP. The mice are rescued completely from heart failure progression.



of links between the cytoskeleton and the initiation of dilated cardiomyopathy. The sarcoglycan-related cardiac myopathies produce two separate lesions, a perivascular defect that creates microfoci of necrosis²⁸, and a widespread loss of myocytes due to plasma-membrane defects elicited during muscle contraction. Mutations in cardiac actin may impair force transmission²³, while defects in nuclear intermediate filaments suggest a role for nuclear signalling and/or instability as causative events²⁹.

Contractility circuits and progression of heart failure

Calcium is the currency of cardiac contractility and relaxation (Fig. 3; for a review, see ref. 14). Cardiac contraction is triggered by the release of Ca^{2+} from stores in the sarcoplasmic reticulum (SR) through a Ca^{2+} -release channel (the ryanodine receptor). Cardiac relaxation is mediated by the re-uptake of calcium by the SR Ca^{2+} pump (Ca^{2+} -ATPase), which maintains SR calcium stores for the next round of cardiac contraction. The β -adrenergic pathway controls this system through the activation of the cyclic AMP (cAMP)-dependent protein kinase A, which phosphorylates an endogenous inhibitor of the calcium pump known as phospholamban³⁰. Phosphorylation of phospholamban prevents its inhibitory effect on the Ca^{2+} -ATPase, leading to an augmentation of pump function and the associated cardiac contractility and relaxation. By coupling miniaturized *in vivo* physiological technology to genetically engineered mouse models³¹, a pivotal role for SR calcium cycling in the progression of heart failure has been discovered³². Inhibiting phospholamban by gene ablation can completely prevent the onset of

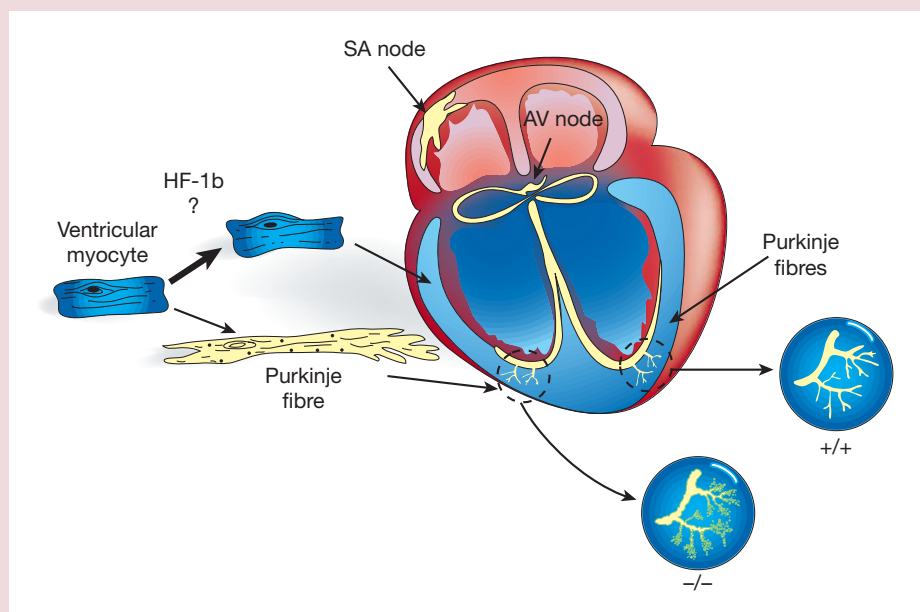
heart failure in mice with the cardiomyopathic MLP mutation³² (Fig. 3). Similarly, somatic gene transfer of the Ca^{2+} -ATPase gene into the myocardium prevents the onset of conserved features of heart failure in acquired models of the disease³³. Therapeutic strategies that aim to activate calcium cycling without augmenting cAMP may prove crucial in avoiding the untoward effects of constitutive increases in cardiac cAMP levels, which can lead to myocyte death and arrhythmogenesis. Intriguingly, the relief of biomechanical stress by unloading the heart via implanting left-ventricular assist devices can temporarily reverse many of the phenotypic features of end-stage heart failure, raising the issue of whether there may be biological pathways that can enhance the reversal of end-stage heart failure. If defects in SR calcium cycling prove to be common to many forms of heart failure, improving calcium cycling by cAMP-independent strategies might serve as new adjunctive therapies in end-stage heart failure.

Phosphorylation of the SR Ca^{2+} -release channel has also been shown to modulate cardiac contractility³⁴. Determining the relative roles of new components of the Ca^{2+} cycling pathway will be critical in identifying the best potential targets for therapeutic drugs. The Ca^{2+} /calmodulin kinase may be of particular interest, as it can directly phosphorylate phospholamban³⁰. Cardiovascular gene databases will enable the identification of new SR components that may have a regulatory or a structural role.

Conduction circuits and arrhythmogenesis

More than 400,000 people die from sudden cardiac death and lethal arrhythmias each year in the United States. High-risk individuals are

Figure 4 A genetic pathway for cardiac sudden death as a result of defects in the transition between ventricular myocytes and conduction system cells (Purkinje fibre). HF-1b may act as a critical switch to activate distinct panels of ion channels and connexins during conduction system formation. A loss of HF-1b results in a 'confused' electrophysiological identity in both ventricular and conduction system cell lineages and conduction defects that lead to cardiac sudden death. SA node, sinoatrial node; AV node, atrioventricular node.



difficult to identify, and no effective therapy for sudden cardiac death exists apart from implantable defibrillators, which have had limited clinical success. Although mutations in ion channels are a rare cause of inheritable cardiac arrhythmias leading to sudden death³⁵, other pathways must account for acquired forms of the disease. As the cells of the electrical conduction system arise from cardiac muscle precursors³⁶, defects in the pathways that guide or maintain the differentiation between ventricular myocyte and Purkinje cell lineages could cause susceptibility to cardiac-associated sudden death (Fig. 4). The clinical association between heart failure and sudden death suggests dysregulation of these signalling pathways during the progression of heart disease.

Direct evidence for this idea has now been obtained from studies of HF-1b, an SP-1-related transcription factor³⁷ that is preferentially expressed in ventricular and conduction system lineages³⁸. Mice deficient in HF-1b survive to term and have normal cardiac structure and function at birth, but display sudden cardiac death and complete penetrance of severe conduction system defects, including spontaneous ventricular tachycardia (accelerated heart rate) and a high-degree atrioventricular heart block³⁸. Continuous electrocardiographic recording documents arrhythmogenesis as the cause of death in these mice, with an identical pattern to that seen in end-stage heart failure. Single-cell analysis indicates an anatomical substrate for arrhythmogenesis that includes a decrease and mislocalization of the connexin 40 and 43 gap junctional proteins, and marked increases in heterogeneity of the action potential. Two independent pathways reveal profound defects in the formation of ventricular Purkinje fibres in these mice. Taken together, these studies identify a new genetic pathway for cardiac-associated sudden death that implicates defects in the transition between ventricular and conduction system cell lineages³⁸. The data point to chaotic impulse propagation in the distal conduction and ventricular cells as the anatomical substrate for cardiac sudden death. Interestingly, mutations in a human cardiac transcription factor, NKX 2.5, are also associated with cardiac conduction defects³⁹. The mouse model should facilitate the use of genomic strategies to identify new downstream target genes in the pathway for cardiac sudden death. A direct evaluation of potential genetic modifiers of cardiac sudden death should now be possible by introducing defined genetic backgrounds into *HF-1b*^{-/-} mutant mice, analogous to the discovery of modifiers of heart failure in cardiomyopathic mice³².

Future perspectives

Genomic databases are ushering in a new era of integrative biology,

providing a universal language for the identification of conserved functions of cardiovascular genes in different species and organ systems. For example, numerous parallels exist between cardiac muscle and neuromuscular disorders, excitability, conduction, terminal differentiation, innervation and neural crest effects. Unsuspected connections between neural and cardiac developmental gene programmes have been found in gene-targeted mice. Conserved myocyte and neuronal cell survival factors have also been discovered, suggesting the existence of yet more. Thus, major new opportunities may lie in wait at the interface of cardiovascular and neuroscience.

To evaluate the real clinical potential of the disease pathways found in mouse models, parallel studies in large animals will be necessary. The difficulty of germ-line gene manipulation in larger animals indicates the value of new investigative strategies that use peptide inhibitors, neutralizing antibodies and somatic gene transfer with smart vectors that allow long-term gene expression. The many quantitatively different endpoints for complex *in vivo* cardiac phenotypes should drive us on to the next frontier that lies at the boundaries of genomic databases, physiology and human disease. We are entering a new renaissance of the integrative biology of the intact organism, and the heart may again play a key role in defining the circuitry for complex *in vivo* physiological traits. □

1. Benson, D. W. *et al.* Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J. Clin. Invest.* **104**, 1567–1573 (1999).
2. Gruber, P. J. *et al.* RXR α deficiency confers genetic susceptibility for aortic sac, conotruncal, atrioventricular cushion, and ventricular muscle defects in mice. *J. Clin. Invest.* **98**, 1332–1343 (1996).
3. Mark, M. *et al.* A genetic dissection of the retinoid signalling pathway in the mouse. *Proc. Nutr. Soc.* **58**, 609–613 (1999).
4. Fleming, A. & Copp, A. J. Embryonic folate metabolism and mouse neural tube defects. *Science* **280**, 2107–2109 (1998).
5. Kreidberg, J. A. *et al.* WT-1 is required for early kidney development. *Cell* **74**, 679–691 (1993).
6. Chen, J., Kubalak, S. W. & Chien, K. R. Ventricular muscle restricted gene targeting of the RXR α gene reveals a non-cell autonomous requirement in cardiac chamber morphogenesis. *Development* **125**, 1943–1949 (1998).
7. Grossfeld, P. D., Rothman, A., Gruber, P. & Chien, K. R. In *Molecular Basis of Cardiovascular Disease: A Companion Text to E. Braunwald's Heart Disease* (ed. Chien, K. R.) 135–165 (Saunders, Cambridge, MA, 1999).
8. Galili, N. *et al.* A region of mouse chromosome 16 is syntenic to the DiGeorge, velocardiofacial syndrome minimal critical region. *Genome Res.* **7**, 17–26 (1997).
9. Yamagishi, H., Garg, V., Matsuoka, R., Thomas, T. & Srivastava, D. A molecular pathway revealing a genetic basis for human cardiac and craniofacial defects. *Science* **283**, 1158–1161 (1999).
10. Lindsay, E. A. *et al.* Congenital heart disease in mice deficient for the DiGeorge syndrome region. *Nature* **401**, 379–383 (1999).
11. Yanagisawa, H. *et al.* Role of endothelin-1/endothelin-A receptor-mediated signaling pathway in aortic arch patterning in mice. *J. Clin. Invest.* **102**, 22–33 (1998).
12. Waldo, K. *et al.* A novel role for cardiac neural crest in heart development. *J. Clin. Invest.* **103**,

- 1499–1507 (1999).
13. Hunter, J. J. & Chien, K. R. Signaling pathways for cardiac hypertrophy and failure. *N. Engl. J. Med.* **341**, 1276–1283 (1999).
14. Chien, K. R. Stress pathways and heart failure. *Cell* **98**, 555–558 (1999).
15. Sheng, Z., Pennica, D., Wood, W. I. & Chien, K. R. Cardiotrophin-1 displays early expression in the murine heart tube and promotes cardiac myocyte survival. *Development* **122**, 419–428 (1996).
16. Hirota, H. *et al.* Loss of a gp130 cardiac muscle cell survival pathway is a critical event in the onset of heart failure during biomechanical stress. *Cell* **97**, 189–198 (1999).
17. Zhao, Y. Y. *et al.* Neuregulins promote survival and growth of cardiac myocytes. Persistence of ErbB2 and ErbB4 expression in neonatal and adult ventricular myocytes. *J. Biol. Chem.* **273**, 10261–10269 (1998).
18. Meyer, D. & Birchmeier, C. Multiple essential functions of neuregulin in development. *Nature* **378**, 386–390 (1995).
19. Slamon, D. *et al.* Addition of Herceptin (humanized anti-Her2 antibody) to first line chemotherapy for Her2 overexpressing mediastinal breast cancer (Her2 + \MBC) markedly increases anti-cancer activity; a randomized, multinational controlled phase III trial. *Proc. Am. Soc. Clin. Oncol.* **17**, 98a (1998).
20. Sadoshima, J. & Izumo, S. The cellular and molecular response of cardiac myocytes to mechanical stress. *Annu. Rev. Physiol.* **59**, 551–571 (1997).
21. Chen, J. & Chien, K. R. Complexity in simplicity: monogenic disorders and complex cardiomyopathies. *J. Clin. Invest.* **103**, 1483–1485 (1999).
22. Arber, S. *et al.* MLP-deficient mice exhibit a disruption of cardiac myofibrillar organization, dilated cardiomyopathy, and heart failure. *Cell* **88**, 393–403 (1997).
23. Olson, T. M., Michels, V. V., Thibodeau, S. N., Tai, Y. S. & Keating, M. T. Actin mutations in dilated cardiomyopathy, a heritable form of heart failure. *Science* **280**, 750–752 (1998).
24. Dalakas, M. C. *et al.* Desmin myopathy, a skeletal myopathy with cardiomyopathy caused by mutations in the desmin gene. *N. Engl. J. Med.* **342**, 770–780 (2000).
25. Satoh, M. *et al.* Structural analysis of the titin gene in hypertrophic cardiomyopathy: identification of a novel disease gene. *Biochem. Biophys. Res. Commun.* **262**, 411–417 (1999).
26. McCoy, G. *et al.* Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease). *Lancet* **355**, 2119–2124 (2000).
27. Dinovi  -Carugo, K., Young, P., Gautel, M. & Saraste, M. Structure of the alpha-actinin rod: molecular basis for cross-linking of actin filaments. *Cell* **98**, 537–546 (1999).
28. Coral-Vazquez, R. *et al.* Disruption of the sarcoglycan-sarcomeric complex in vascular smooth muscle: a novel mechanism for cardiomyopathy and muscular dystrophy. *Cell* **98**, 465–474 (1999).
29. Bonne, G. *et al.* Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nature Genet.* **21**, 285–288 (1999).
30. Tada, M. & Toyofuku, T. Molecular regulation of phospholamban function and expression. *Trends Cardiovasc. Med.* **8**, 330–340 (1998).
31. Chien, K. R. Genes and physiology: molecular physiology in genetically engineered animals. *J. Clin. Invest.* **97**, 901–909 (1996).
32. Minamisawa, S. *et al.* Chronic phospholamban-sarcoplasmic reticulum calcium ATPase interaction is the critical calcium cycling defect in dilated cardiomyopathy. *Cell* **99**, 313–322 (1999).
33. Miyamoto, M. E. *et al.* Adenoviral gene transfer of SERCA2a improves left ventricular function in aortic-banded rats in transition to heart failure. *Proc. Natl Acad. Sci. USA* **97**, 793–798 (2000).
34. Marx, S. O. *et al.* PKA phosphorylation dissociates FKBP12.6 from the calcium release channel (ryanodine receptor): defective regulation in failing hearts. *Cell* **101**, 365–376 (2000).
35. Curran, M. E., Sanguinetti, M. C. & Keating, M. T. in *Molecular Basis of Cardiovascular Disease: A Companion Text to E. Braunwald's Heart Disease* (ed. Chien, K. R.) 302–311 (Saunders, Cambridge, MA, 1999).
36. Gourdie, R. G., Kubalak, S. & Mikawa, T. Conducting the embryonic heart: orchestrating development of specialized cardiac tissues. *Trends Cardiovasc. Med.* **9**, 18–26 (1999).
37. Zhu, H. *et al.* A novel, tissue restricted zinc finger protein (HF-1b) binds to the cardiac regulatory element (HF-1b/MEF-2) in the rat myosin light chain-2 gene. *Mol. Cell Biol.* **13**, 4432–4444 (1993).
38. Nguyen-Tran, V. T. B. *et al.* A novel genetic pathway for cardiac sudden death via defects in the transition between ventricular and conduction system cell lineages. *Cell* **102**, 671–682 (2000).
39. Schott, J. J. *et al.* Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* **281**, 108–111 (1998).
40. Franz, T. Persistent truncus arteriosus in the Splotch mutant mouse. *Anat. Embryol.* **180**, 457–464 (1989).
41. Clouthier, D. E. *et al.* Cranial and cardiac neural crest defects in endothelin-A receptor-deficient mice. *Development* **125**, 813–824 (1998).
42. Kurihara, Y. *et al.* Elevated blood pressure and craniofacial abnormalities in mice deficient in endothelin-1. *Nature* **368**, 703–710 (1994).
43. Yanagisawa, H. *et al.* Dual genetic pathways of endothelin-mediated intercellular signaling revealed by targeted disruption of endothelin converting enzyme-1 gene. *Development* **125**, 825–836 (1998).
44. Ya, J. *et al.* Sox4-deficiency syndrome in mice is an animal model for common trunk. *Circ. Res.* **83**, 986–994 (1998).
45. Reaume, A. G. *et al.* Cardiac malformation in neonatal mice lacking connexin 43. *Science* **267**, 1831–1834 (1995).
46. Donovan, M. J., Hahn, R., Tessarollo, L. & Hempstead, B. L. Identification of an essential nonneuronal function of neurotrophin 3 in mammalian cardiac development. *Nature Genet.* **14**, 210–213 (1996).
47. Tessarollo, L. *et al.* Targeted deletion of all isoforms of the trkC gene suggests the use of alternate receptors by its ligand neurotrophin-3 in neuronal development and implicates trkC in normal cardiogenesis. *Proc. Natl Acad. Sci. USA* **94**, 14776–14781 (1997).
48. Brannan, C. I. *et al.* Targeted disruption of the neurofibromatosis type-1 gene leads to developmental abnormalities in heart and various neural crest-derived tissues. *Genes Dev.* **8**, 1019–1029 (1994).
49. Tullio, A. N. *et al.* Nonmuscle myosin II-B is required for normal development of the mouse heart. *Proc. Natl Acad. Sci. USA* **94**, 12407–12412 (1997).
50. Sanford, L. P. *et al.* TGF  2 knockout mice have multiple developmental defects that are nonoverlapping with other TGF   knockout phenotypes. *Development* **124**, 2659–2670 (1997).
51. Srivastava, D. *et al.* Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nature Genet.* **16**, 154–160 (1997).
52. Iida, K. *et al.* Essential roles of the winged helix transcription factor MFH-1 in aortic arch patterning and skeletogenesis. *Development* **124**, 4627–4638 (1997).
53. Takihara, Y. *et al.* Targeted disruption of the mouse homologue of the *Drosophila* polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects. *Development* **124**, 3673–3682 (1997).
54. Emanuel, B. S., Budarf, M. L. & Scambler, P. J. in *Heart Development* (eds Harvey, R. P. & Rosenthal, N.) 463–478 (Academic, San Diego, 1999).

Acknowledgements

I thank T. Nakamura, M. Hoshijima and S. Minamisawa for assistance with the figures, and J. Chen, S. Evans and K. Knowlton for helpful discussions. The author's laboratory is supported by grants from NIH, the Jean LeDucq Foundation, and an Endowed Chair from the American Heart Association (California Affiliate).

Atherosclerosis

Aldons J. Lusis

Department of Medicine, Department of Microbiology, Immunology and Molecular Genetics, Department of Human Genetics, and Molecular Biology Institute, University of California, Los Angeles, California 90095, USA (e-mail: jlusis@mednet.ucla.edu)

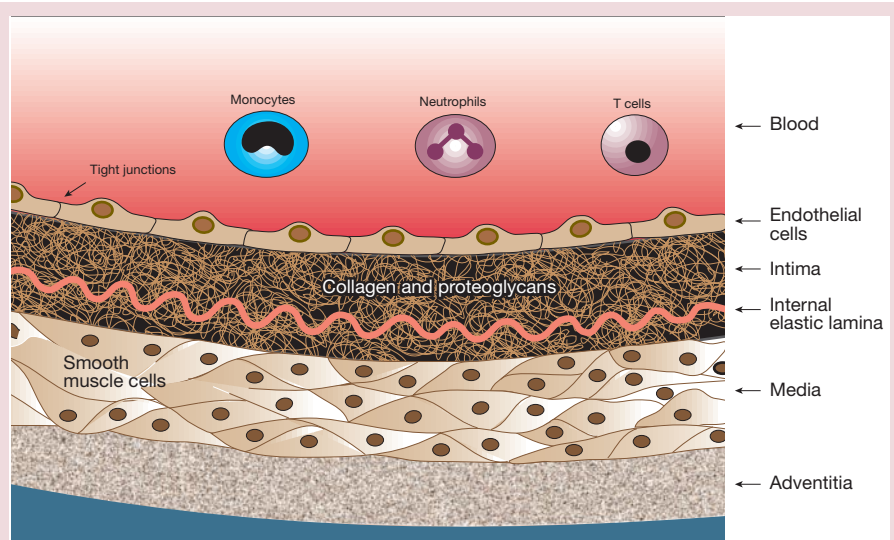
Atherosclerosis, a disease of the large arteries, is the primary cause of heart disease and stroke. In westernized societies, it is the underlying cause of about 50% of all deaths. Epidemiological studies have revealed several important environmental and genetic risk factors associated with atherosclerosis. Progress in defining the cellular and molecular interactions involved, however, has been hindered by the disease's aetiological complexity. Over the past decade, the availability of new investigative tools, including genetically modified mouse models of disease, has resulted in a clearer understanding of the molecular mechanisms that connect altered cholesterol metabolism and other risk factors to the development of atherosclerotic plaque. It is now clear that atherosclerosis is not simply an inevitable degenerative consequence of ageing, but rather a chronic inflammatory condition that can be converted into an acute clinical event by plaque rupture and thrombosis.

Atherosclerosis is a progressive disease characterized by the accumulation of lipids and fibrous elements in the large arteries. The anatomy of a normal artery is shown in Fig. 1. The early lesions of atherosclerosis consist of subendothelial accumulations of cholesterol-engorged macrophages, called 'foam cells'. In humans, such 'fatty streak' lesions can usually be found in the aorta in the first decade of life, the coronary arteries in the second decade, and the cerebral arteries in the third or fourth decades. Because of differences in blood flow dynamics, there are preferred sites of lesion formation within the arteries. Fatty streaks are not clinically significant, but they are the precursors of more advanced lesions characterized by the accumulation of lipid-rich necrotic debris and smooth muscle cells (SMCs). Such 'fibrous lesions' typically have a 'fibrous cap' consisting of SMCs and extracellular matrix that encloses a lipid-rich 'necrotic core'. Plaques can become increasingly complex, with calcification, ulceration at the luminal surface, and haemorrhage from small vessels that grow into the lesion from the media of the blood vessel wall. Although advanced lesions can grow sufficiently large to block blood flow, the most important clinical complication is an acute occlusion due to the

formation of a thrombus or blood clot, resulting in myocardial infarction or stroke. Usually, the thrombosis is associated with rupture or erosion of the lesion.

The events of atherosclerosis have been greatly clarified by studies in animal models, including rabbits, pigs, non-human primates and rodents. Mice deficient in apolipoprotein E (apoE) or the low-density lipoprotein (LDL) receptor develop advanced lesions and are the models most used in genetic and physiological studies¹. Figure 2 shows stages in the development of atherosclerotic plaques in experimental animals. The first observable change in the artery wall following the feeding of a high-fat, high-cholesterol diet is the accumulation of lipoprotein particles and their aggregates in the intima at sites of lesion predilection (Fig. 2a, b). Within days or weeks, monocytes can be observed adhering to the surface of the endothelium. The monocytes then transmigrate across the endothelial monolayer into the intima, where they proliferate, differentiate into macrophages and take up the lipoproteins, forming foam cells (Fig. 2c, d)². With time, the foam cells die, contributing their lipid-filled contents to the necrotic core of the lesion. Some fatty streaks subsequently accumulate SMCs, which migrate from the medial layer. With the secretion of

Figure 1 Structure of a normal large artery. A large artery consists of three morphologically distinct layers. The intima, the innermost layer, is bounded by a monolayer of endothelial cells on the luminal side and a sheet of elastic fibres, the internal elastic lamina, on the peripheral side. The normal intima is a very thin region (size exaggerated in this figure) and consists of extracellular connective tissue matrix, primarily proteoglycans and collagen. The media, the middle layer, consists of SMCs. The adventitia, the outer layer, consists of connective tissues with interspersed fibroblasts and SMCs.



fibrous elements by the smooth muscle cells, occlusive fibrous plaques develop and increase in size. Initially, the lesions grow towards the adventitia until a critical point is reached, after which they begin to expand outwards and encroach on the lumen. The lesions continue to grow by the migration of new mononuclear cells from the blood, which enter at the shoulder of the vessel; this is accompanied by cell proliferation, extracellular matrix production and the accumulation of extracellular lipid (Fig. 2e). Atherogenesis can be viewed as a 'response to injury', with lipoproteins or other risk factors as the injurious agents^{2,3}.

A very complex aetiology

Epidemiological studies over the past 50 years have revealed numerous risk factors for atherosclerosis (Table 1). These can be grouped into factors with an important genetic component, and those that are largely environmental. The relative abundance of the different plasma lipoproteins appears to be of primary importance, as raised levels of atherogenic lipoproteins are a prerequisite for most forms of the disease. With the exception of gender, and the level of lipoprotein(a), each of the genetic risk factors involves multiple genes. This complexity can be clearly observed in genetic crosses in animals maintained under similar environmental conditions; such studies in rodents have revealed dozens of genetic loci that contribute to lipoprotein levels, body fat and other risk factors⁴. Another level of complexity involves the interactions between risk factors. Frequently, these are not simply additive; for example, the effects of hypertension on coronary heart disease (CHD) are considerably amplified if cholesterol levels are high⁵.

The importance of genetics and environment in human CHD has been examined in many family and twin studies⁶. Within a population, the heritability of atherosclerosis (the fraction of disease explained by genetics) has been high in most studies, frequently exceeding 50%. Population migration studies, on the other hand, clearly show that the environment explains much of the variation in disease incidence between populations. Thus, the common forms of CHD result from the combination of an unhealthy environment, genetic susceptibility and our increased lifespan⁵.

Cellular and molecular interactions

Pathological studies have revealed a defined series of changes in the vessel during atherogenesis (Fig. 2) and showed that blood-derived inflammatory cells, particularly monocytes/macrophages, have a key role. Tissue culture studies with vascular cells and monocytes/macrophages suggested possible pathways of disease initiation and progression. They provided evidence for the central role of the endothelium in mediating inflammation, and suggested that accumulation of oxidatively modified LDL in the intima contributes significantly to monocyte recruitment and foam-cell formation. During the past decade, understanding of the molecular mechanisms in atherogenesis has been revolutionized by studies in transgenic and gene-targeted mice⁷. These have allowed *in vivo* testing of hypotheses, although it should be noted that studies in mice are limited by significant species differences compared with humans, and that reliable mouse models for thrombosis involving lesion rupture have not been developed.

Lesion initiation

The endothelium, with its intercellular tight junctional complexes, functions as a selectively permeable barrier between blood and tissues. It has both sensory and executive functions, and can generate effector molecules that regulate thrombosis, inflammation, vascular tone and vascular remodelling. For example, removal of the endothelium results in a burst of SMC migration and proliferation, which subsides when the endothelium regenerates⁸. Among the important physical forces acting on endothelial cells (ECs) is fluid shear stress, which has effects on EC morphology. Cells in the tubular regions of arteries, where blood flow is uniform and laminar, are ellipsoid in shape and aligned in the direction of flow. Cells in regions of arterial branching or curvature, where flow is disturbed, have polygonal shapes and no particular orientation. These latter areas show increased permeability to macromolecules such as LDL and are preferential sites for lesion formation⁸.

As shown in Fig. 3, a primary initiating event in atherosclerosis is the accumulation of LDL in the subendothelial matrix. Accumulation is greater when levels of circulating LDL are raised, and both the transport and retention of LDL are increased in the preferred sites for

Table 1 Genetic and environmental factors associated with atherosclerosis and coronary heart disease (CHD)

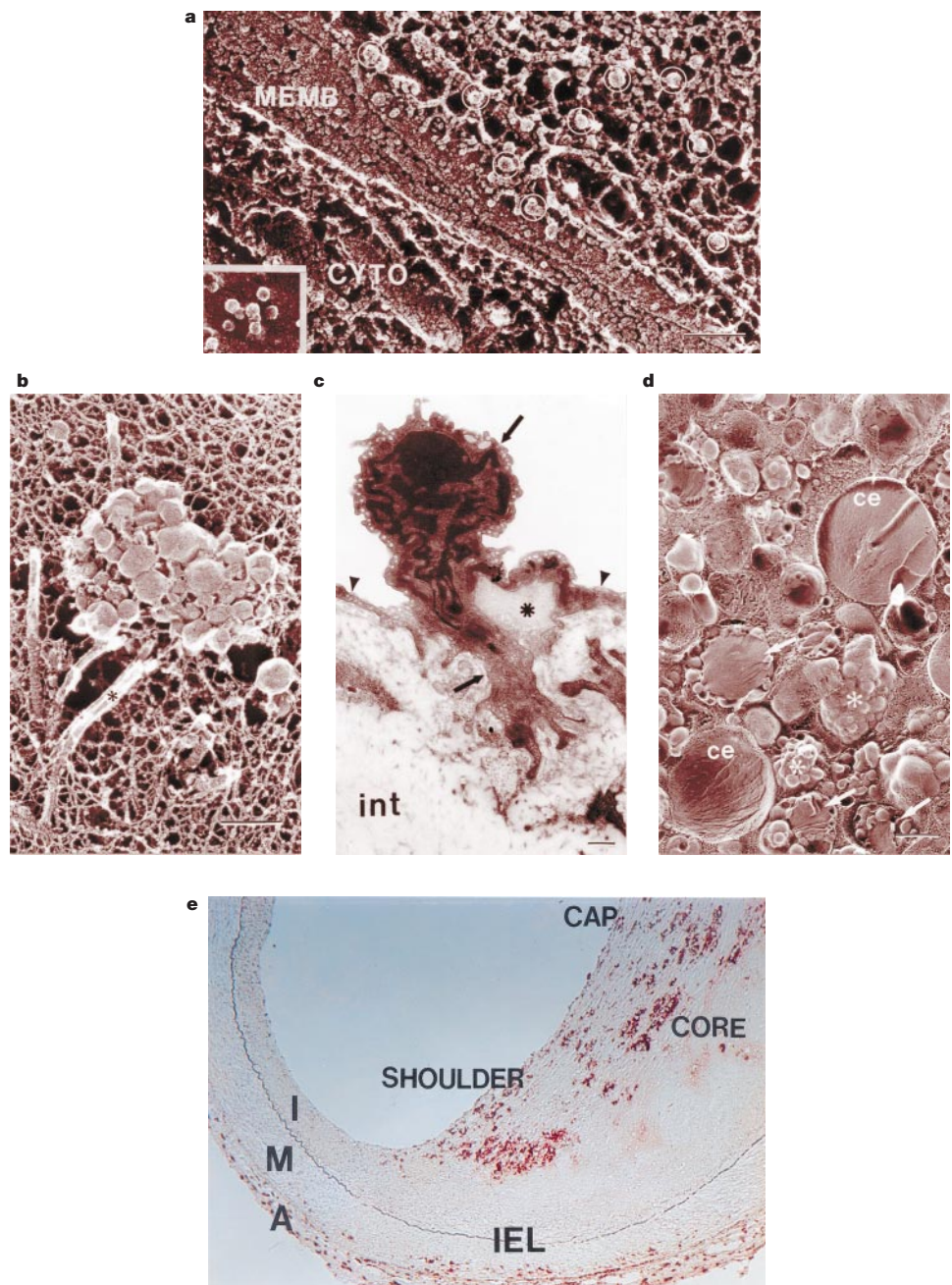
Factors with a strong genetic component

Elevated levels of LDL/VLDL	Associations demonstrated in epidemiological studies and supported by studies of genetic disorders and animal models. Clinical trials have shown benefits of cholesterol reduction ⁵⁴ .
Reduced levels of HDL	Associations demonstrated by numerous epidemiological studies and supported by studies of genetic diseases and animal models ⁵⁸ .
Elevated levels of lipoprotein(a)	Associations observed in many, but not all, epidemiological studies. Animal studies have been contradictory ⁵⁹ .
Elevated blood pressure	Associations observed in epidemiological studies. Clinical trials have demonstrated benefits of blood pressure reduction, with particularly strong effects on stroke ^{54,60} .
Elevated levels of homocysteine	Associations have been observed in epidemiological studies, and homocystinuria results in severe occlusive vascular disease ⁶⁰ .
Family history	When all known risk factors are controlled for, family history remains a very significant independent factor ⁶ .
Diabetes and obesity	Associations observed in epidemiological studies and in studies with animal models ⁵⁴ .
Elevated levels of haemostatic factors	Significant independent associations have been observed with elevated levels of fibrinogen, plasminogen activator inhibitor type 1 and platelet reactivity ⁵⁴ .
Depression and other behavioural traits	Associations observed in several population studies ⁶¹ .
Gender (male)	Below age 60, men develop CHD at more than twice the rate of women ⁵⁸ .
Systemic inflammation	Elevated levels of inflammatory molecules such as C-reactive protein are associated with CHD, as are inflammatory diseases such as rheumatoid arthritis ⁶² .
Metabolic syndrome	This cluster of metabolic disturbances, with insulin resistance as a central feature, is strongly associated with CHD ⁵ .

Environmental factors

High-fat diet	Population migration and epidemiological studies indicate strong associations with lifestyle, and diet appears to be the most significant factor. High-fat, high-cholesterol diets are usually required for development of atherosclerosis in experimental animals ⁵⁴ .
Smoking	Strong associations observed in numerous epidemiological studies. Clinical trials have demonstrated the benefit of stopping smoking ⁵⁴ .
Low antioxidant levels	Results of clinical trials with antioxidants have not been conclusive. Fat-soluble antioxidants protect against atherosclerosis in experimental animals, however ⁶³ .
Lack of exercise	Significant independent associations with CHD ⁵⁴ .
Infectious agents	Epidemiological studies provide suggestive evidence for associations with various infectious agents, such as <i>Chlamydia pneumoniae</i> . Preliminary animal studies support the relationship ⁶⁴ .

Figure 2 Stages in the development of atherosclerotic plaques. **a**, In the first stages, lipoprotein is trapped in the subendothelial matrix. The freeze-etch electron micrograph shows the accumulation of 23-nm LDL particles (circled) in the matrix of a rabbit atrial-ventricular valve following incubation with LDL (inset). An endothelial cell at lower left shows the plasma membrane (MEMB) and cytoplasm (CYTO)⁷¹. Magnification $\times 141,372$; scale bar, 0.1 μm . **b**, Lipoprotein aggregation is seen in this freeze-etch electron micrograph of rabbit intima following administration of a bolus of LDL. The aggregated particles are surrounded by matrix and collagen fibrils (asterisk)⁷². Magnification $\times 52,876$; scale bar, 0.2 μm . **c**, Monocyte transmigration. The thin-section electron micrograph of a cross-section of the aorta of a 9-week-old apoE-deficient mouse shows a monocyte (arrow) moving between two endothelial cells (arrowheads) to enter the intima (int). The asterisk denotes a cluster of lipid underneath the endothelial cell¹. Magnification $\times 10,078$; scale bar, 0.5 μm . **d**, Foam-cell formation. Freeze-etch electron micrograph of the cytoplasm of a macrophage foam cell in the intima of a rabbit fed a high-fat diet for two weeks. Large lipid droplets with the onion skin configuration typical of cholesterol esters (ce) as well as other lipid-filled compartments (arrows) can be recognized. Some compartments contain large aggregated LDL particles (asterisk) resembling those in **b**. Magnification $\times 21,542$; scale bar, 0.5 μm . **e**, Fibrous lesion. Light micrograph ($\times 400$) of a section of an advanced human coronary atherosclerotic lesion that has been immunostained for the macrophage-specific antigen EMB-11 (red). A, adventitia; I, intima; IEL, internal elastic lamina; M, media. Photographs courtesy of A. Mottino, J. Frank and T. Drake, UCLA.



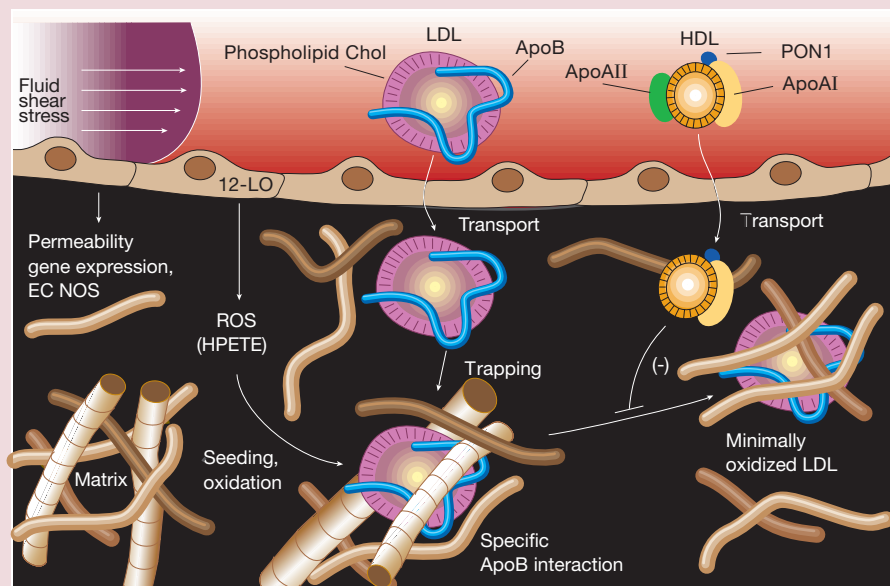
lesion formation. LDL diffuses passively through EC junctions, and its retention in the vessel wall seems to involve interactions between the LDL constituent apolipoprotein B (apoB) and matrix proteoglycans⁹. In addition to LDL, other apoB-containing lipoproteins, namely lipoprotein(a) and remnants, can accumulate in the intima and promote atherosclerosis. Lipoprotein(a), a particle resembling LDL but containing an additional polypeptide termed apolipoprotein(a) that is linked to apoB by a disulphide bridge, seems to be particularly atherogenic owing to its additional effects on fibrinolysis and SMC growth¹⁰.

Native LDL is not taken up by macrophages rapidly enough to generate foam cells, and so it was proposed that LDL is somehow 'modified' in the vessel wall¹¹. It has subsequently been shown that trapped LDL does indeed undergo modification, including oxidation, lipolysis, proteolysis and aggregation, and that such

modifications contribute to inflammation as well as to foam-cell formation. One of the modifications most significant for early lesion formation is lipid oxidation as a result of exposure to the oxidative waste of vascular cells. Such modifications initially give rise to 'minimally oxidized' LDL species that have pro-inflammatory activity but may not be sufficiently modified to be recognized by macrophage scavenger receptors. Mice lacking 12/15-lipoxygenase have considerably diminished atherosclerosis, suggesting that this enzyme may be an important source of reactive oxygen species in LDL oxidation¹². Lipoxygenases insert molecular oxygen into polyenoic fatty acids, producing molecules such as hydroperoxyeicosatetraenoic acid (HPETE), which are likely to be transferred across the cell membrane to 'seed' the extracellular LDL.

High-density lipoprotein (HDL) is strongly protective against atherosclerosis. An important mechanism underlying this protective

Figure 3 Lesion initiation. Sites of lesion predilection are determined in part by haemodynamic forces acting on endothelial cells. These influence the permeability of the endothelial barrier and expression of endothelial cell (EC) genes such as that for nitric oxide synthase (NOS). An important initiating event is the retention of LDL and other apolipoprotein B (apoB)-containing lipoproteins as a result of interaction with matrix components. The LDL undergoes oxidative modification as a result of interaction with reactive oxygen species (ROS) including products of 12/15 lipoxygenase (12-LO) such as HPETE. Oxidation of LDL is inhibited by HDL, which contains the antioxidant protein serum paraoxonase (PON1).



effect is the role of HDL in the removal of excess cholesterol from peripheral tissues. But in addition, HDL also protects by inhibiting lipoprotein oxidation. The antioxidant properties of HDL are due in part to serum paraoxonase, an esterase carried on HDL that can degrade certain biologically active oxidized phospholipids^{13,14}.

Inflammation

Atherosclerosis is characterized by the recruitment of monocytes and lymphocytes, but not neutrophils, to the artery wall (Fig. 4). A triggering event for this process is the accumulation of minimally oxidized LDL, which stimulates the overlying ECs to produce a number of pro-inflammatory molecules, including adhesion molecules and growth factors such as macrophage colony-stimulating factor (M-CSF). The biological activity of minimally oxidized LDL is contained primarily in its phospholipid fraction, and three active oxidation products resulting from the scission or rearrangement of unsaturated fatty acids have been identified¹⁵. Oxidized LDL can also inhibit the production of nitric oxide (NO), a chemical mediator with multiple anti-atherogenic properties, including vasorelaxation. Mice lacking endothelial NO synthase showed enhanced atherosclerosis, due in part to raised blood pressure¹⁶. In addition to oxidized

LDL, a number of other factors are likely to modulate inflammation, including haemodynamic forces, homocysteine levels, sex hormones, and infection. Diabetes may promote inflammation in part by the formation of advanced endproducts of glycation that interact with endothelial receptors¹⁷.

The entry of particular types of leukocytes into the artery wall is mediated by adhesion molecules and chemotactic factors. After cultured ECs are exposed to oxidized LDL, they will bind monocytes but not neutrophils. The first step in adhesion, the 'rolling' of leukocytes along the endothelial surface, is mediated by selectins which bind to carbohydrate ligands on leukocytes. Studies of mice deficient in P- and E-selectins or the cell adhesion molecule ICAM, revealed the role of these adhesion molecules in atherosclerosis^{18,19}. The firm adhesion of monocytes and T cells to endothelium can be mediated by the integrin VLA-4 on these cells, which interacts with both VCAM-1 on the endothelium and the CS-1 splice variant of fibronectin. Both *in vitro* and *in vivo* studies suggested that these interactions have a role in atherosclerosis²⁰. Finally, mice deficient in monocyte chemotactic protein (MCP-1) or its receptor CCR2 had significantly reduced atherosclerotic lesions, suggesting that

Figure 4 Inflammation. Minimally oxidized LDL stimulates the overlying endothelial cells to produce adhesion molecules, chemotactic proteins such as monocyte chemotactic protein-1 (MCP-1), and growth factors such as macrophage colony-stimulating factor (M-CSF), resulting in the recruitment of monocytes to the vessel wall. Oxidized LDL has other effects, such as inhibiting the production of NO, an important mediator of vasodilation and expression of endothelial leukocyte adhesion molecules (ELAMs). Among endothelial cell adhesion molecules likely to be important in the recruitment of leukocytes are ICAM-1, P-selectin, E-selectin, PCAM-1 and VCAM-1. Important adhesion molecules on monocytes include β 2 integrin, VLA-4, and PCAM-1. Advanced glycosylation endproducts (AGEs) are formed in diabetes and these promote inflammation via specific receptors on endothelial cells.

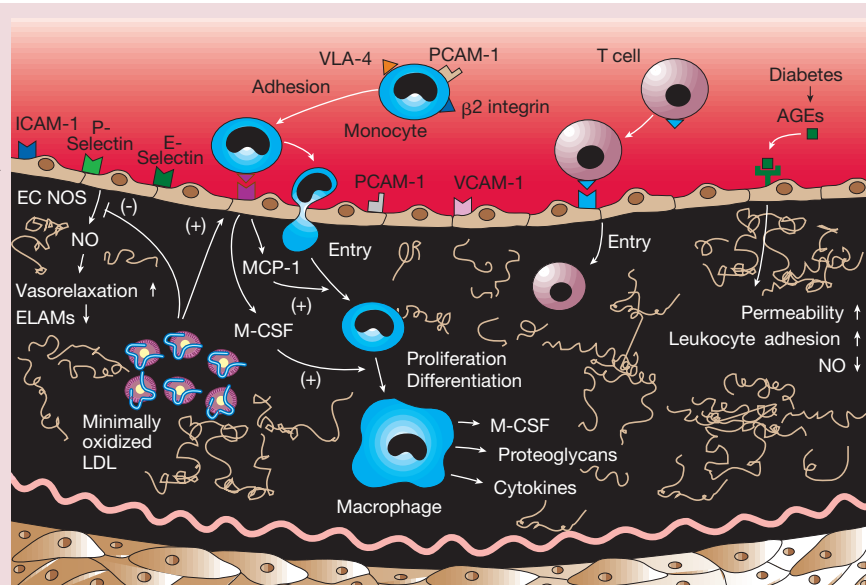
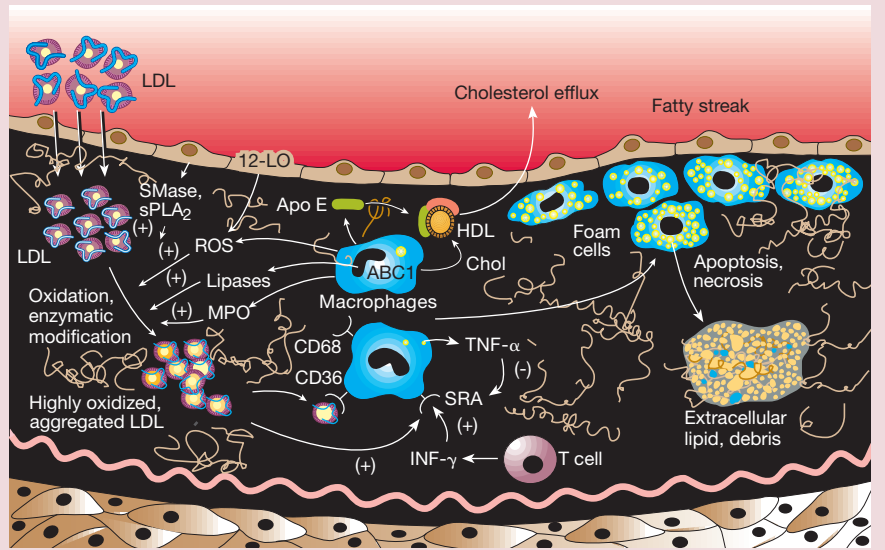


Figure 5 Foam-cell formation. Highly oxidized aggregated LDL is formed in the vessel as a result of the action of reactive oxygen species (ROS) and the enzymes sphingomyelinase (SMase), secretory phospholipase 2 (sPLA₂), other lipases, and myeloperoxidase (MPO). The oxidized aggregated LDL is recognized by macrophage scavenger receptors such as SR-A, CD36 and CD68. Scavenger receptor expression is mediated by cytokines such as tumour necrosis factor- α (TNF- α) and interferon- γ (IFN- γ). Foam cells secrete apolipoprotein E (apoE), which may facilitate removal of excess cellular cholesterol. The death of foam cells leaves behind a growing mass of extracellular lipids and other cell debris.



MCP-1/CCR2 interaction has a role in monocyte recruitment in atherosclerosis^{21,22}.

The cytokine M-CSF stimulates the proliferation and differentiation of macrophages, and influences various macrophage functions such as expression of scavenger receptors. Mice with a spontaneous null mutation of M-CSF had dramatically reduced lesions, suggesting an obligatory role for macrophages in lesion formation²³.

Foam-cell formation

LDL must be extensively modified ('highly oxidized') before it can be taken up sufficiently rapidly by macrophages to form foam cells (Fig. 5). This modification presumably involves reactive oxygen species produced by ECs and macrophages, but several enzymes are also thought to be involved, including myeloperoxidase, sphingomyelinase and a secretory phospholipase, all of which occur in human atherosclerotic lesions. Myeloperoxidase generates highly reactive species such as hypochlorous acid and tyrosyl radical, and myeloperoxidase-modified LDL binds to macrophage scavenger receptors²⁴. Sphingomyelinase may promote lipoprotein aggregation, leading to increased retention and enhanced uptake by macrophages²⁵. Finally, a secretory phospholipase (group II sPLA₂) can promote LDL oxidation, and transgenic mice overexpressing the enzyme show increased atherosclerosis²⁶.

The rapid uptake of highly oxidized (and otherwise modified) LDL particles by macrophages, leading to foam-cell formation, is

mediated by a group of receptors that recognize a wide array of ligands. Two such 'scavenger' receptors, SR-A and CD36, appear to be of primary importance, and mice lacking either receptor show a modest reduction in atherosclerotic lesions^{27,28}. The expression of scavenger receptors is regulated by peroxisome proliferator-activated receptor- γ , a transcription factor whose ligands include oxidized fatty acids, and by cytokines such as tumour necrosis factor- α and interferon- γ (IFN- γ)²⁹.

Macrophages actively secrete apoE, and this may promote cholesterol efflux to HDL, thereby inhibiting the transformation of macrophages to foam cells. Evidence for this role of apoE comes from bone marrow transplantation studies showing that mice transplanted with marrow from apoE-null mice develop much larger lesions than mice receiving marrow from control mice³⁰. Interestingly, mice deficient in ACAT1, the enzyme responsible for cholesterol esterification in macrophages, are still able to develop significant lesions³¹.

Fibrous plaques

Fibrous plaques are characterized by a growing mass of extracellular lipid, mostly cholesterol and its ester, and by the accumulation of SMCs and SMC-derived extracellular matrix (Fig. 6). Cytokines and growth factors secreted by macrophages and T cells are important for SMC migration and proliferation and extracellular matrix production.

Recent studies have shown that the interaction of CD40 with its ligand CD40L (CD154) makes an important contribution to the

Figure 6 Formation of fibrous plaques. A number of risk factors, including elevated levels of homocysteine and angiotensin II (produced through the action of angiotensin-converting enzyme, ACE), stimulate the migration or proliferation of SMCs. Oestrogens exert beneficial effects on plasma lipoprotein levels and they also stimulate production of NO and prostacyclin by endothelial cells. The interaction of CD40 and CD40 ligand (CD40L) stimulates T lymphocytes (T cells) and macrophages to express cytokines such as IFN- γ that can influence inflammation, SMC growth and matrix accumulation. The intimal SMCs secrete extracellular matrix and give rise to a fibrous cap.

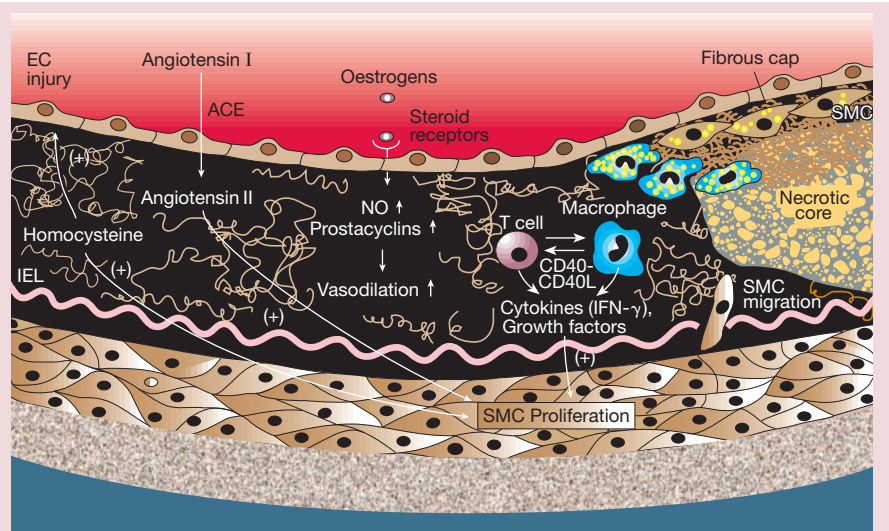
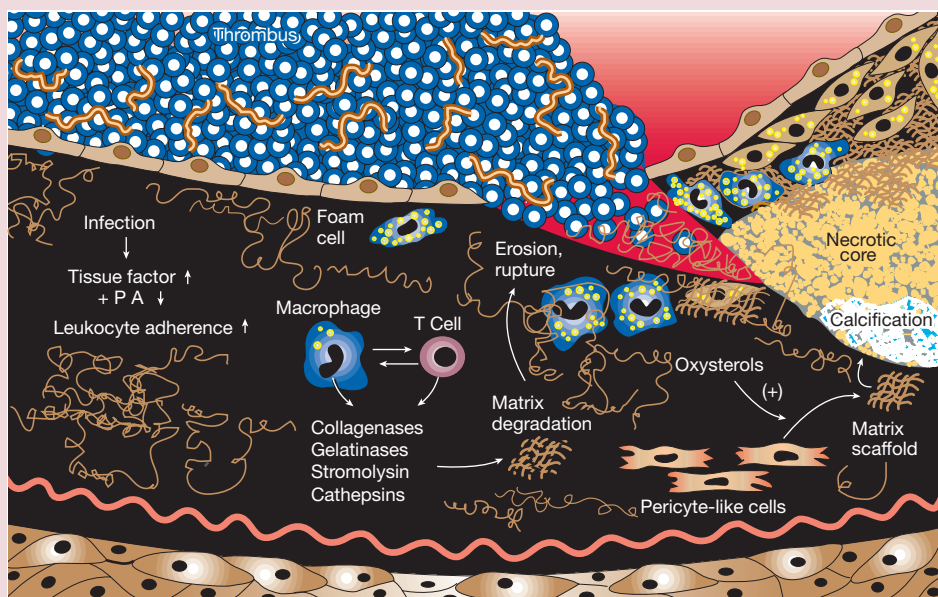


Figure 7 Complex lesions and thrombosis.

Vulnerable plaques with thin fibrous caps result from degradation of matrix by various proteinases such as collagenases, gelatinases, stromolysin and cathepsins and by inhibition of matrix secretion. Among various factors that may destabilize plaques and promote thrombosis are infection, which may have systemic effects such as induction of acute phase proteins and local effects such as increased expression of tissue factor and decreased expression of plasminogen activator (PA). The calcification of lesions appears to be an active, regulated process involving the secretion by pericyte-like cells in the intima of a scaffold for calcium phosphate deposition. The formation of a thrombus, consisting of adherent platelets and fibrin crosslinks, usually results from plaque rupture, exposing tissue factor in the necrotic core.



development of advanced lesions³². This interaction was first recognized as being essential to major immune reactions involving T and B cells, but it is now clear that CD40 is also expressed on macrophages, ECs and SMCs. The engagement of CD40 and CD40L results in the production of inflammatory cytokines, matrix-degrading proteases and adhesion molecules. Studies using CD40L-null mice or neutralizing antibodies to CD40L have shown that disruption of the interaction results in smaller lesions that are less inflammatory and more fibrous³². Although studies with immunodeficient mice originally indicated a modest role of lymphocytes in atherogenesis³³, studies of CD40–CD40L³², of antibodies to oxidized LDL epitopes³⁴, and of the T-lymphocyte product IFN- γ ³⁵ are consistent with a major role for lymphocytes.

Several risk factors seem to contribute to the development of fibrous lesions, including elevated homocysteine, hypertension and hormones. Elevated homocysteine levels appear to injure ECs and to stimulate proliferation of vascular SMCs³⁶. Some of the effects of raised blood pressure on atherosclerosis seem to be mediated by components of the renin–angiotensin pathway. For example, angiotensin II directly stimulates SMC growth and the production of extracellular matrix. Studies with spontaneously hypertensive rats (SHR) indicate that raised blood pressure stimulates expression of platelet-derived growth factor, a potent mitogen for SMCs³⁷. Oestrogen has multiple anti-atherogenic properties, including effects on plasma lipoprotein levels and stimulation of prostacyclin and NO production³⁸.

Infection by cytomegalovirus has been linked to atherosclerosis and arterial restenosis (a narrowing of the vessel lumen due to vascular remodelling following angioplasty)³⁵. On the basis of *in vitro* studies, a plausible mechanism for this link is stimulation of SMC migration by the virus-coded chemokine receptor US28 (ref. 39). Cytomegalovirus infection is also associated with inactivation of the p53 protein, and p53-null mice exhibited increased SMC proliferation and accelerated atherosclerosis⁴⁰.

The monoclonal patchiness of atherosclerotic lesions originally suggested that the disease may involve a nonmalignant transformation of SMCs, but this patchiness has now been shown to result from normal development⁴¹. Nevertheless, evidence consistent with oncogene activation, loss of heterozygosity and microsatellite instability in human lesions has been reported⁵.

Advanced lesions and thrombosis

Pathological studies suggest that the development of thrombus-mediated acute coronary events depends principally on the

composition and vulnerability of a plaque rather than the severity of stenosis (Fig. 7). Vulnerable plaques generally have thin fibrous caps and increased numbers of inflammatory cells. Maintenance of the fibrous cap reflects matrix production and degradation, and products of inflammatory cells are likely to influence both processes. For example, T cells produce IFN- γ , which inhibits the production of matrix by SMCs, and macrophages produce various proteases that degrade extracellular matrix, including interstitial collagenase, gelatinases and stromolysin³. Rupture frequently occurs at the lesion edges, which are rich in foam cells, suggesting that factors contributing to inflammation may also influence thrombosis. In this regard, it is notable that the incidence of myocardial infarction and stroke increases during acute infections.

The stability of atherosclerotic lesions may also be influenced by calcification and neovascularization, common features of advanced lesions. Intimal calcification is an active process in which pericyte-like cells secrete a matrix scaffold which subsequently becomes calcified, akin to bone formation. The process is regulated by oxysterols and cytokines⁴². The growth of small vessels from the media may provide a conduit for entry of inflammatory cells⁴³.

The thrombogenicity of the lesion core is likely to depend on the presence of tissue factor, a key protein in the initiation of the coagulation cascade. The production of tissue factor by ECs and macrophages is enhanced by oxidized LDL, infection or the ligation of CD40 on ECs to CD40L on inflammatory cells⁴⁴. The expression of other molecules mediating thrombosis, such as plasminogen activator, may also be important.

Genetic dissection of atherosclerosis

Although the common forms of atherosclerosis are multifactorial, studies of rare mendelian forms have provided the most important insights into the disease (Table 2). Studies of familial hypercholesterolaemia helped unravel the pathways that regulate plasma cholesterol metabolism, knowledge of which was important for the development of cholesterol-lowering drugs. In the past year, Tangier disease, a rare recessive disorder characterized by the virtual absence of circulating HDL, was shown to be due to mutations in the gene for the ATP-binding-cassette (ABC) transporter 1, providing an excellent candidate cause for more common forms of HDL deficiency^{45,46}. Recently found mutations in the mineralocorticoid receptor, a kidney protein that is involved in the body's handling of salt, explain why some women have a sharp rise in blood pressure during pregnancy⁴⁷.

Table 2 Mendelian disorders relevant to atherosclerosis

Trait	Disease (gene)	Characteristics
Elevated LDL/VLDL levels	Familial hypercholesterolaemia (LDL receptor) ⁵	Dominant disorder characterized by very high LDL-cholesterol levels and early CHD
	Familial defective apoB-100 (apoB) ⁵	Dominant disorder due to apoB mutations that affect binding to LDL receptor; less severe than FH
Low HDL levels	ApoAI deficiency (apoAI) ⁵	In the homozygous state, null mutations of apoAI result in the virtual absence of HDL and early CHD
	Tangier disease (ABC1 transporter) ^{45,46}	This recessive disorder results in the inability of cells to export cholesterol and phospholipids, resulting in very low levels of HDL
Coagulation	Various genetic disorders of haemostasis ⁵	Unlike rare disorders of lipid metabolism where atherosclerotic disease is a primary manifestation, genetic disorders of haemostasis usually present either as increased risk of bleeding or thrombosis (usually venous), with no outstanding effect on atherogenesis
Elevated homocysteine	Homocystinuria (cystathionine β -synthetase) ⁴⁶	Recessive metabolic disorder resulting in very high levels of homocysteine and severe occlusive vascular disease
Diabetes, type 2	MODY1 (hepatocyte nuclear factor 4 α) ⁵ , MODY2 (glucokinase) ⁵ , MODY3 (hepatocyte nuclear factor 1 α) ⁵	MODY1, 2, and 3 are characterized by the development of non-insulin dependent diabetes mellitus in young adults
Hypertension	Glucocorticoid-remediable aldosteronism (hybrid gene from crossover of 11- β -hydroxylase and aldosterone synthase) ⁶⁰	Dominant disorder with early-onset hypertension and stroke
	Liddle's syndrome (epithelial sodium channel) ⁶⁰	Dominant disorder with hypertension and metabolic alkalosis
	Mineralocorticoid receptor ⁴⁷	Early-onset hypertension associated with pregnancy

In contrast to the mendelian disorders, attempts to identify genes for the common, complex forms of atherosclerosis have met with mixed success. Studies of candidate genes have revealed a number that show significant or suggestive association or linkage with traits relevant to atherosclerosis, but our understanding remains incomplete (Table 3). Large-scale sequencing is now underway to identify polymorphisms for many other candidate genes for hypertension, diabetes and other traits relevant to atherosclerosis⁴⁸. In an attempt to identify atherosclerosis genes, whole-genome scans for loci associated with diabetes, hyperlipidaemia, low HDL levels and hypertension have been performed⁴⁹. But few loci with significant evidence of linkage have been found, emphasizing the complexity of these traits.

The use of animal models is a potentially powerful way of identifying genes that contribute to common forms of atherosclerosis. Mice and rats—the most useful mammals for genetic studies—have common variations in many traits relevant to atherosclerosis, and orthologous genes frequently contribute to a trait in rodents and humans⁵⁰. Mapping and identification of genes contributing to complex traits is easier in rodents than in humans, as shown by the recent identification of a diabetes gene in the SHR rat model⁵¹. Studies in animal models should be particularly useful for the identification of genetic factors influencing vascular cell functions; for example, differences in susceptibility to atherosclerosis between certain strains of mice seem to be due to variation that affects EC responses to oxidized LDL⁵². During this decade it is likely that genome-wide approaches, such as expression array studies and large-scale animal mutagenesis studies, will become widely used in atherosclerosis research.

As a result of the genome projects and large-scale sequencing, tens of thousands of single-nucleotide polymorphisms are being identified and a catalogue of all common variations in humans will be generated over the next few years. This raises the possibility of whole-genome association studies. Given the rapid development of DNA chip technology, it should be possible to type large numbers of polymorphisms in many thousands of individuals. There are, however, significant unresolved issues involving linkage disequilibrium and statistical analysis⁵³ in this approach.

New therapies

Effective drugs for lowering cholesterol and high blood pressure have been developed. In particular, the statins lower levels of atherogenic lipoproteins and dramatically decrease clinical events and mortality from atherosclerosis⁵⁴. Nevertheless, heart disease and stroke remain by far the most common causes of death in westernized societies, and

Table 3 Common genetic variations contributing to CHD and its risk factors

Trait	Gene	Variation
LDL/VLDL	ApoE ⁵	Three common missense alleles explain ~5% of variance in cholesterol levels
HDL levels	Hepatic lipase ⁶⁵	Promoter polymorphism
	ApoAI-CIII-AIV cluster ⁶⁵	Multiple polymorphisms
	Cholesteryl ester transfer protein ⁵	Common null mutations (Japanese); missense polymorphisms
	Lipoprotein lipase ⁶⁶	Missense polymorphisms
Lipoprotein(a)	Apolipoprotein(a) ⁶⁹	Many alleles explain >90% variance
Homocysteine	Methylene tetrahydrofolate reductase ⁵	Missense polymorphism
Coagulation	Fibrinogen B ⁵	Promoter polymorphism
	Plasminogen activator inhibitor type 1 ⁵	Promoter polymorphism
	Factor VIII ⁵	Missense polymorphism
Blood pressure	Angiotensinogen ⁶⁰	Missense and promoter polymorphisms
	β_2 -adrenergic receptor ⁶⁰	Missense polymorphism
	Alpha-adducin ⁶⁰	Missense polymorphism
CHD	Angiotensin-converting enzyme ⁶⁷	Insertion-deletion polymorphism
	Serum paraoxonase ^{13,14}	Missense polymorphism affecting enzymatic activity
	Haemachromatosis-associated gene ⁶⁸	Missense polymorphism
	Endothelial nitric oxide synthase ⁶⁹	Missense polymorphism
	Factor XIII ⁷⁰	Missense polymorphism

Only genes exhibiting evidence of linkage or association in two or more studies are cited.

new weapons, particularly agents that block disease at the level of the vessel wall or that raise anti-atherogenic HDL, are needed.

Over the past decade, a number of promising new targets have been identified, as discussed above and shown in Figs 3–7. For example, interruption of the CD40–CD40L system may have clinical benefits for plaque stability³². The identification of the ABC transporter presents exciting new opportunities for treatment of low HDL



levels. It has also become clear that HDLs are functionally very heterogeneous⁵⁵. Thus, rather than attempting to increase levels of HDL, it may be more productive to focus on functional properties such as its antioxidant activity. Preliminary studies in animals suggest that it may be possible not only to block the development of atherosclerosis but also to achieve significant regression⁵⁶. The most critical clinical aspect of atherosclerosis is plaque rupture and thrombosis. Although useful mouse models for this have not been developed, a transgenic hypertensive and hyperlipidaemic rat model showed evidence of myocardial infarction⁵⁷.

Diagnosis and risk assessment

Catheterization is the gold standard for diagnosis of atherosclerosis, but it is expensive and carries significant risk. Reliable noninvasive methods of diagnosis are urgently needed. Certain biochemical markers for the disease, such as C-reactive protein, and some noninvasive procedures, such as extravascular ultrasound and ultrafast computerized tomography, should prove useful but have limitations.

As our understanding of the genetics of atherosclerosis increases, genetic diagnosis will become increasingly important. The anticipated 'biallelic map' of the genome is likely to drive the evolution of new technologies for gene screening, from high-throughput, genome-wide methods to testing for particular gene variants in individuals. One application of screening will be to distinguish different forms of the disease so that pharmacological intervention can be better targeted. Atherosclerosis is heterogeneous, and the most appropriate therapy will depend on the particular variety of disease. Classification is already used clinically, as patients are grouped according to the variety of risk factors they display, but genetic testing should greatly expand the subdivisions of the disease.

Another potential benefit of genetic studies is testing for susceptibility. Because CHD and stroke are disorders of adults, knowledge of a propensity to disease could be available many years before clinical disease develops, permitting early intervention. Testing for LDL, HDL and blood pressure have long been advocated as a way of identifying individuals at increased risk, and other factors have emerged more recently as risk indicators (Table 1). Once the genes contributing to common forms of the disease have been identified, along with the particular mutations involved, DNA-based tests may add greatly to our ability to assess risk. But given the importance of environmental influences and the complex genetic aetiology of atherosclerosis, efficient screening procedures are unlikely to be available in the near future. □

- Tamminen, M., Mottino, G., Qiao, J. H., Breslow, J. L. & Frank, J. S. Ultrastructure of early lipid accumulation in apoE-deficient mice. *Arterioscl. Thromb. Vasc. Biol.* **19**, 847–853 (1999).
- Ross, R. The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature* **362**, 801–809 (1993).
- Libby, P. Changing concepts of atherogenesis. *J. Intern. Med.* **247**, 349–358 (1999).
- Mehrabian, M., Wen, P.-Z., Fislir, J., Davis, R. C. & Lusis, A. J. Genetic loci controlling body fat, lipoprotein metabolism, and insulin levels in a multifactorial mouse model. *J. Clin. Invest.* **101**, 2485–2496 (1998).
- Lusis, A. J., Weinreb, A. & Drake, T. A. in *Textbook of Cardiovascular Medicine* (ed. Topol, E. J.) 2389–2413 (Lippincott-Raven, Philadelphia, 1998).
- Goldbourt, U. & Neufeld, H. N. Genetic aspects of arteriosclerosis. *Arteriosclerosis* **6**, 357–377 (1988).
- Smithies, O. & Maeda, N. Gene targeting approaches to complex diseases: atherosclerosis and essential hypertension. *Proc. Natl Acad. Sci. USA* **92**, 5266–5272 (1995).
- Gimbrone, M. A. Jr Vascular endothelium, hemodynamic forces, and atherogenesis. *Am. J. Pathol.* **155**, 1–5 (1999).
- Boren, J. *et al.* Identification of the principal proteoglycan-binding site in LDL. A single-point mutation in apo-B100 severely affects proteoglycan interaction without affecting LDL receptor binding. *J. Clin. Invest.* **101**, 2658–2664 (1998).
- Grainger, D. J., Kemp, P. R., Liu, A. C., Lawn, R. M. & Metcalfe, J. C. Activation of transforming growth factor- β is inhibited in transgenic apolipoprotein(a) mice. *Nature* **370**, 460–462 (1994).
- Goldstein, J. L., Ho, Y. K., Basu, S. K. & Brown, M. S. Binding sites on macrophages that mediate uptake and degradation of acetylated low density lipoprotein, producing massive cholesterol deposition. *Proc. Natl Acad. Sci. USA* **76**, 333–337 (1979).
- Cyrus, T. *et al.* Disruption of 12/15-lipoxygenase diminishes atherosclerosis in apoE-deficient mice. *J. Clin. Invest.* **103**, 1597–1604 (1999).
- Hegele, R. A. Paraoxonase—genes and disease. *Ann. Med.* **31**, 217–224 (1999).
- Shih, D. M. *et al.* Combined serum paraoxonase knockout/apolipoprotein E knockout mice exhibit increased lipoprotein oxidation and atherosclerosis. *J. Biol. Chem.* **276**, 17527–17535 (2000).
- Watson, A. D. *et al.* Structural identification by mass spectrometry of oxidized phospholipids in minimally oxidized low density lipoprotein that induce monocyte/endothelial interactions and evidence for their presence *in vivo*. *J. Biol. Chem.* **272**, 13597–13607 (1997).
- Knowles, J. W. *et al.* Enhanced atherosclerosis and kidney dysfunction in eNOS(–/–) apoE(–/–) mice are ameliorated by enalapril treatment. *J. Clin. Invest.* **105**, 451–458 (2000).
- Hofmann, M. A. *et al.* RAGE mediates a novel proinflammatory axis: a central surface receptor for S100/calgranulin polypeptides. *Cell* **97**, 889–901 (1999).
- Dong, Z. M. *et al.* The combined role of P- and E-selectins in atherosclerosis. *J. Clin. Invest.* **102**, 145–152 (1998).
- Collins, R. G. *et al.* P-selectin or intercellular adhesion molecule (ICAM-1) deficiency substantially protects against atherosclerosis in apolipoprotein E-deficient mice. *J. Exp. Med.* **191**, 189–194 (2000).
- Shih, P. T. *et al.* Blocking very late antigen-4 integrin decreases leukocyte entry and fatty streak formation in mice fed an atherogenic diet. *Circ. Res.* **84**, 345–351 (1998).
- Gu, L. *et al.* Absence of monocyte chemoattractant protein-1 reduces atherosclerosis in low density lipoprotein-deficient mice. *Mol. Cell* **2**, 275–281 (1998).
- Boring, L., Gosling, J., Cleary, M. & Charo, I. F. Decreased lesion formation in CCR2–/– mice reveals a role for chemokines in the initiation of atherosclerosis. *Nature* **394**, 894–897 (1998).
- Smith, J. D. *et al.* Decreased atherosclerosis in mice deficient in both macrophage colony-stimulating factor (op) and apolipoprotein E. *Proc. Natl Acad. Sci. USA* **92**, 8264–8268 (1995).
- Podrez, E. A. *et al.* Macrophage scavenger receptor CD36 is the major receptor for LDL modified by monocyte-generated reactive nitrogen species. *J. Clin. Invest.* **105**, 1095–1108 (2000).
- Marathe, S., Kuriakose, G., Williams, K. J. & Tabas, I. Sphingomyelinase, an enzyme implicated in atherogenesis, is present in atherosclerotic lesions and binds to specific components of the subendothelial extracellular matrix. *Arterioscl. Thromb. Vasc. Biol.* **19**, 2648–2658 (1999).
- Ivandic, B. *et al.* Role of group II secretory phospholipase A₂ in atherosclerosis I. Increased atherogenesis and altered lipoproteins in transgenic mice expressing group IIa phospholipase A₂. *Arterioscl. Thromb. Vasc. Biol.* **19**, 1284–1290 (1999).
- Suzuki, H. *et al.* A role for macrophage scavenger receptors in atherosclerosis and susceptibility to infection. *Nature* **386**, 292–296 (1997).
- Febbraio, M. *et al.* Targeted disruption of the class B scavenger receptor CD36 protects against atherosclerosis lesion development in mice. *J. Clin. Invest.* **105**, 1049–1056 (2000).
- Tontonoz, P., Nagy, L., Alvarez, J. L., Thomazy, V. A. & Evans, R. M. PPAR gamma promotes monocyte/macrophage differentiation and uptake of oxidized LDL. *Cell* **93**, 241–252 (1998).
- Fazio, S. *et al.* Increased atherogenesis in mice reconstituted with apolipoprotein E null macrophages. *Proc. Natl Acad. Sci. USA* **94**, 4647–4652 (1997).
- Accad, M. *et al.* Massive xanthomatosis and altered composition of atherosclerotic lesions in hyperlipidemic mice lacking acyl CoA:cholesterol acyltransferase 1. *J. Clin. Invest.* **105**, 711–719 (2000).
- Schönbeck, U., Sukhova, G. K., Shimizu, K., Mach, F. & Libby, P. Inhibition of CD40 signaling limits evolution of established atherosclerosis in mice. *Proc. Natl Acad. Sci. USA* **97**, 7458–7463 (2000).
- Fyfe, A. L., Qiao, J. H. & Lusis, A. J. Immune deficient mice develop typical atherosclerotic fatty streaks when fed an atherogenic diet. *J. Clin. Invest.* **94**, 2516–2520 (1994).
- Shaw, P. X. *et al.* Natural antibodies with T15 idiotype may act in atherosclerosis apoptotic clearance and protective immunity. *J. Clin. Invest.* **105**, 1731–1740 (2000).
- Gupta, S. *et al.* IFN- γ potentiates atherosclerosis in apoE knock-out mice. *J. Clin. Invest.* **99**, 2752–2761 (1997).
- Gerhard, G. T. & Duell, P. B. Homocysteine and atherosclerosis. *Curr. Opin. Lipidol.* **10**, 417–429 (1999).
- Negoro, N. *et al.* Blood pressure regulates platelet-derived growth factor A-chain gene expression in vascular smooth muscle cells *in vivo*. An autocrine mechanism promoting hypertensive vascular hypertrophy. *J. Clin. Invest.* **95**, 1140–1150 (1995).
- Nathan, L. & Chaudhuri, G. Estrogens and atherosclerosis. *Annu. Rev. Pharmacol. Toxicol.* **37**, 477–515 (1997).
- Strebblow, D. N. *et al.* The human cytomegalovirus chemokine receptor US28 mediates vascular smooth muscle cell migration. *Cell* **99**, 511–520 (1999).
- Guevara, N. V., Kim, H.-S., Antonova, E. L. & Chan, L. The absence of p53 accelerates atherosclerosis by increasing cell proliferation *in vivo*. *Nature Med.* **5**, 335–339 (1999).
- Schwartz, S. M. & Murray, C. E. Proliferation and the monoclonal origin of atherosclerotic lesions. *Annu. Rev. Med.* **49**, 437–460 (1998).
- Watson, K. E. *et al.* TGF- β 1 and 25-hydroxycholesterol stimulate osteoblast-like vascular cells to calcify. *J. Clin. Invest.* **93**, 2106–2113 (1994).
- Moulton, K. S. & Folkman, J. in *Molecular Basis of Cardiovascular Disease* (ed. Chien, K. R.) 393–410 (Saunders, Philadelphia, 1999).
- Schönbeck, U. *et al.* CD40 ligation induces tissue factor expression in human vascular smooth muscle cells. *Am. J. Pathol.* **156**, 7–14 (2000).
- Young, S. G. & Fielding, C. J. The ABCs of cholesterol efflux. *Nature Genet.* **22**, 316–318 (1999).
- Orsö, E. *et al.* Transport of lipids from Golgi to plasma membrane is defective in Tangier disease patients and ABC1-deficient mice. *Nature Genet.* **24**, 192–196 (2000).
- Geller, D. S. *et al.* Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. *Science* **289**, 119–123 (2000).
- Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* **22**, 231–238 (1999).
- Krushkal, J. *et al.* Genome-wide linkage analyses of systolic blood pressure using highly discordant siblings. *Circulation* **99**, 1407–1410 (1999).
- Stoll, M. *et al.* New target regions for human hypertension via comparative genomics. *Genome Res.* **10**, 473–482 (2000).
- Aitman, T. J. *et al.* Identification of CD36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nature Genet.* **21**, 76–83 (1999).
- Shi, W., Haberland, M. E., Jien, M. L., Shih, D. M. & Lusis, A. J. Endothelial responses to oxidized lipoproteins determine genetic susceptibility to atherosclerosis in mice. *Circulation* **102**, 75–81 (2000).
- Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847–856 (2000).
- Assmann, G., Cullen, P., Jossa, F., Lewis, B. & Mancini, M. Coronary heart disease: reducing the risk. *Arterioscl. Thromb. Vasc. Biol.* **19**, 1819–1824 (1999).
- Navab, M. *et al.* The yin and the yang of oxidation in the development of the fatty streak. A review based on the 1994 George Lyman Duff Memorial Lecture. *Arterioscl. Thromb. Vasc. Biol.* **16**, 831–842 (1996).
- Desumont, C. *et al.* Complete atherosclerosis regression after human apoE gene transfer in apoE deficient/nude mice. *Arterioscl. Thromb. Vasc. Biol.* **20**, 435–442 (2000).

57. Herrera, V. L. *et al.* Spontaneous combined hyperlipidemia, coronary heart disease and decreased survival in Dahl salt-sensitive hypertensive rats transgenic for human cholesteryl ester transfer protein. *Nature Med.* **5**, 1383–1389 (1999).
58. Gordon, D. J. & Rifkind, B. M. High-density lipoprotein—the clinical implications of recent studies. *N. Engl. J. Med.* **321**, 1311–1316 (1989).
59. Kronenberg, F. *et al.* Role of lipoprotein(a) and apolipoprotein(a) phenotype in atherogenesis. *Circulation* **100**, 1154–1160 (1999).
60. Luft, F. C. Molecular genetics of human hypertension. *J. Hypertens.* **16**, 1871–1878 (1998).
61. Glassman, A. H. & Shapiro, P. A. Depression and the course of coronary artery disease. *Am. J. Psychiatry* **155**, 4–11 (1998).
62. Kugiyama, K. *et al.* Circulating levels of secretory type II phospholipase A₂ predict coronary events in patients with coronary artery disease. *Circulation* **100**, 1280–1284 (1999).
63. Steinberg, D. & Witztum, J. L. in *Molecular Basis of Cardiovascular Disease* (ed. Chien, K. R.) 458–475 (Saunders, Philadelphia, 1999).
64. Hu, H., Pierce, G. N. & Zhong, G. The atherogenic effects of chlamydia are dependent on serum cholesterol and specific to *Chlamydia pneumoniae*. *J. Clin. Invest.* **103**, 747–753 (1999).
65. Cohen, J. C., Wang, Z., Grundy, S. M., Stoesz, M. R. & Guerra, R. Variation at the hepatic lipase and apolipoprotein AI/CIII/AIV loci is a major cause of genetically determined variation in plasma HDL cholesterol levels. *J. Clin. Invest.* **94**, 2377–2384 (1994).
66. Wittrup, H. H., Tybjaerg-Hansen, A. & Nordestgaard, B. G. Lipoprotein lipase mutations, plasma lipids and lipoproteins, and risk of ischemic heart disease: a meta-analysis. *Circulation* **99**, 2901–2907 (1999).
67. Samani, N. J., Thompson, J. R., O'Toole, L., Channer, K. & Woods, K. L. A meta-analysis of the association of the deletion allele of the angiotensin-converting enzyme gene with myocardial infarction. *Circulation* **94**, 708–712 (1996).
68. Tuomainen, T.-P. *et al.* Increased risk of acute myocardial infarction in carriers of the hemochromatosis gene Cys282 Tyr mutation. *Circulation* **100**, 1274–1279 (1999).
69. Hingorani, A. D. *et al.* A common variant of the endothelial nitric oxide synthase (Glu298 → Asp) is a major risk factor for coronary artery disease in the UK. *Circulation* **100**, 1515–1520 (1999).
70. Franco, R. F. *et al.* Factor XIII and the risk of myocardial infarction. *Haematologica* **85**, 67–71 (2000).
71. Niveststein-Post, P., Mottino, G., Fogelman, A. & Frank, J. An ultrastructural study of lipoprotein accumulation in cardiac valves of the rabbit. *Arterioscl. Thromb. Vasc. Biol.* **14**, 1151–1161 (1994).
72. Niveststein, P. F., Fogelman, A. M., Mottino, G. & Frank, J. S. Lipid accumulation in rabbit aorta intima 2 hours after bolus infusion of low density lipoprotein. A deep-etch and immunolocalization study of ultrarapidly frozen tissue. *Arterioscl. Thromb. Vasc. Biol.* **11**, 1795–1805 (1991).

Acknowledgements

I thank R. Chen and K. Wong for help with the preparation of this manuscript and L. Olson for help with the illustrations. Work in my laboratory was supported by NIH grants.



Vascular-specific growth factors and blood vessel formation

George D. Yancopoulos, Samuel Davis, Nicholas W. Gale, John S. Rudge, Stanley J. Wiegand & Jocelyn Holash

Regeneron Pharmaceuticals, Inc., 777 Old Saw Mill River Road, Tarrytown, New York 10591, USA

A recent explosion in newly discovered vascular growth factors has coincided with exploitation of powerful new genetic approaches for studying vascular development. An emerging rule is that all of these factors must be used in perfect harmony to form functional vessels. These new findings also demand re-evaluation of therapeutic efforts aimed at regulating blood vessel growth in ischaemia, cancer and other pathological settings.

Until recently, vascular endothelial growth factor (VEGF) was the only growth factor proven to be specific and critical for blood vessel formation^{1–3}. Other long-known factors, such as the fibroblast growth factors (FGFs), had profound effects in various endothelial cell assays⁴. But such factors were also known to be nonspecific in that they could act on many other cell types, and it was questionable whether the assays used to evaluate them were physiologically relevant. For example, the most widely used assays involved adding putative angiogenic agents to cornea pocket models, or to chick chorioallantoic membranes^{5,6}. In such assays, FGFs could robustly induce new vessel growth, but there was limited ability to evaluate the induced vessels functionally, or to determine the relevance of these inductions for normal vascular development.

A recent explosion of newly discovered growth factors acting on the vascular endothelium has coincided with application of powerful new genetic approaches to the problem of vascular development^{7,8}. The vascular endothelium-specific growth factors now include five members of the VEGF family, four members of the angiopoietin family, and at least one member of the large ephrin family (Fig. 1). For almost all of these and their receptors, mouse models involving genetic disruption and/or transgenic misexpression have contributed to an understanding of their normal physiological roles, as well as of their pathological capabilities. A rule that is emerging is that all of these factors must be used in perfect harmony, in a complementary and coordinated manner, to form functional vessels⁷. In addition, many other growth factors that are not vascular endothelium-specific are also required for blood vessel formation, such as members of the platelet-derived growth factor or transforming growth factor- β families, although these factors also have critical roles for many other systems as well^{8–10}. Furthermore, there are myriad other gene products — ranging from transcription factors to members of the Notch family — that have been shown crucial for vessel formation⁸. In an attempt to do justice to the topic, this review will focus only on the vascular endothelium-specific growth factors, and how they are involved in vessel formation.

The recent explosion in identifying and characterizing physiological regulators of blood vessel growth demands re-evaluation of therapeutic efforts aimed at regulating blood vessel growth — whether it be promoting vascular ingrowth to replenish ischaemic tissue, blocking vessel growth in order to blunt tumours, or repairing damaged and leaky

vessels during inflammation or other pathological settings. The privilege of hindsight makes some of the bold, early therapeutic efforts directed towards ischaemic disease, based on random delivery of a single growth factor to grow an entirely new functional network of vessels, now appear somewhat naive and even misguided. On the other hand, recent insights continue to support the notion that blockade of even a single growth factor might limit disease-induced vascular growth, with the most compelling evidence supporting approaches based on blockade of VEGF. Furthermore, recent advances indicate previously unanticipated clinical applications for vascular growth factors, such as the use of angiopoietin-1 (Ang1) for the repair of damaged and leaky vessels.

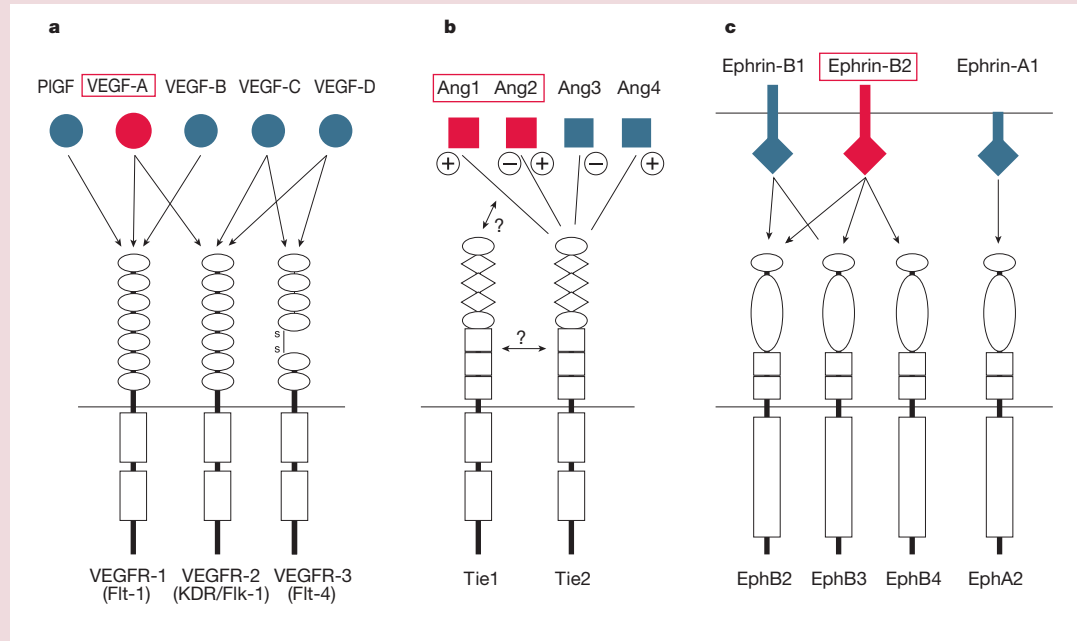
Vasculogenesis and angiogenic remodelling

Vessel formation can occur by a number of different processes⁴. Early in development, vessel formation occurs by a process referred to as vasculogenesis (Fig. 2, stage A), in which endothelial cells differentiate and proliferate *in situ* within a previously avascular tissue, and then coalesce to form a primitive tubular network. This primary network includes some of the major vessels in the embryo, such as the aorta and major veins, as well as a honeycomb-like plexus connecting these major vessels. Angiogenic remodelling refers to the process by which this initial network is modified — through both pruning and vessel enlargement — to form the interconnecting branching patterns characteristic of the mature vasculature (Fig. 2, stage B). During this time, vessel walls also mature, as endothelial cells integrate tightly with supporting cells (such as smooth muscle cells and pericytes) and surrounding matrix (Fig. 2, stage C). A different process, referred to as angiogenic sprouting, involves the sprouting from existing vessels into a previously avascular tissue. In some cases, it seems as if mature vessels must first be destabilized to allow for subsequent sprouting (Fig. 2, stages D, F); once again, vessels formed by sprouting are initially immature and must further develop. Angiogenic sprouting is responsible for vascularizing certain structures during normal development, such as the neural tube or the retina, and for most new vessel formation in the adult. Destabilization of vessels can also apparently lead to vascular regression (Fig. 2, stage E), as described below.

Emerging model of vascular formation

Recent insights have led to a model of vascular formation that attempts to incorporate the known vascular-specific growth

Figure 1 Schematic representation of three families of vascular growth factors and their receptor interactions. **a**, VEGFs; **b**, angiopoietins; **c**, ephrins. The four factors that are discussed in detail in this review are highlighted in red. In **b**, '+' or '-' indicates whether the particular angiopoietin activates or blocks the Tie2 receptor, whereas '?' indicates that a potential interaction has not yet been confirmed experimentally. In **c**, only those members of the large ephrin ligand family (and only their counterpart Eph receptors) that have been implicated in vascular growth are shown.



factors^{7,11–14}, and the details of this model will be a major subject of this review. According to this model, the first characterized vascular-specific growth factor, VEGF, maintains its position as the most critical driver of vascular formation, as it is required to initiate the formation of immature vessels by vasculogenesis or angiogenic sprouting (Fig. 2, stages A, F), during development as well as in the adult. Ang1 and ephrin-B2 are subsequently required for further remodelling and maturation of this initially immature vasculature (Fig. 2, stages B, C), with ephrin-B2 being particularly important in distinguishing developing arterial and venous vessels, as will be discussed in more detail below.

Following vessel maturation, Ang1 seems to continue to be important in maintaining the quiescence and stability of the mature vasculature (Fig. 2, stage C). Disruption of this stabilizing signal coincides with reinitiation of vascular remodelling in the adult — as occurs in the adult female reproductive system or in tumours (Fig. 2, stage D, and see below). Such de-stabilization seems to involve the autocrine induction — by the endothelium to be remodelled — of a natural antagonist of Ang1, termed Ang2 (Fig. 2, stage D). VEGFs, angiopoietins and ephrin-B2 apparently recapitulate their developmental roles during vascular remodelling in the adult, and administration of individual factors to the adult allows them to reprise these roles but not to trigger the entire process (see below). Thus VEGF administration can initiate vessel formation in adult animals, but by itself promotes formation of only leaky, immature and unstable vessels. In contrast, Ang1 administration seemingly further stabilizes and protects the adult vasculature, making it resistant to the damage and leak induced by VEGF or inflammatory challenges. Altogether, it is becoming clear that precise understanding of the normal developmental roles of the VEGFs, the angiopoietins and the ephrins will greatly aid in understanding how to manipulate these growth factor systems for therapeutic benefit.

VEGF, its relatives, and their receptors

VEGF was initially defined, characterized and purified for its ability to induce vascular leak and permeability, as well as for its ability to promote vascular endothelial cell proliferation^{1,2}. Thus, it was originally termed vascular permeability factor as well as VEGF. Although most research efforts have focused on its growth-promoting ability, recent findings are once again highlighting its potent permeability-inducing effects, and in particular their role in disease. Other members of the VEGF family were identified based on their homology to VEGF³. The various members of the VEGF family have

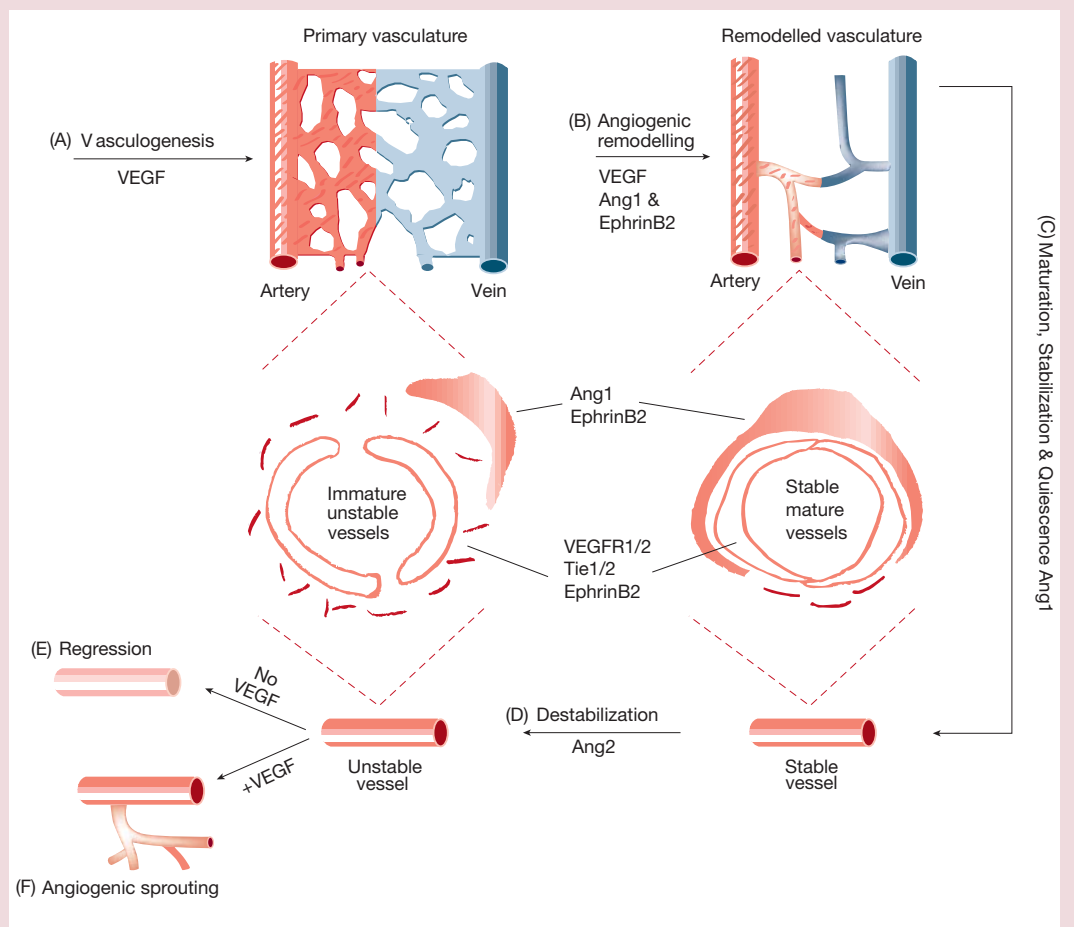
overlapping abilities to interact with a set of cell-surface receptors³ that trigger responses to these factors (Fig. 1a). The main receptors that seem to be involved in initiating signal transduction cascades in response to the VEGFs comprise a family of closely related receptor tyrosine kinases consisting of three members now termed VEGFR-1 (previously known as Flt-1), VEGFR-2 (previously known as KDR or Flk-1) and VEGFR-3 (previously known as Flt-3). In addition, there are a number of accessory receptors such as the neuropilins¹⁵ which seem to be involved primarily in modulating binding to the main receptors, although roles in signalling have not been ruled out.

VEGFR-2 seems to mediate the major growth and permeability actions of VEGF, whereas VEGFR-1 may have a negative role, either by acting as a decoy receptor or by suppressing signalling through VEGFR-2. Thus, mice engineered to lack VEGFR-2 fail to develop a vasculature and have very few endothelial cells¹⁶, whereas mice lacking VEGFR-1 seem to have excess formation of endothelial cells which abnormally coalesce into disorganized tubules¹⁷. Mice engineered to express only a truncated form of VEGFR-1, lacking its kinase domain, appear rather normal, consistent with the notion that the primary role of VEGFR-1 may be that of a decoy receptor¹⁸. VEGFR-3 may be important during blood vessel development, but is most unique based on its expression on lymphatic vessels, for which it seems to be critical¹⁹. The first VEGF relative identified is known as placental growth factor (PIGF), and until recently little was known about its normal function, in part because mice engineered to lack PIGF were overtly normal^{8,20}. Recent findings indicate that adult mice lacking PIGF exhibit deficiencies in certain models of adult vascular remodelling, raising the interesting possibility that the activity of PIGF may be limited to these settings⁸. VEGF-C, based on its ability to bind the lymphatic-specific VEGFR-3, seems to be important for lymphatic development, and transgenic overexpression of VEGF-C leads to lymphatic hyperplasia²¹. Mice lacking VEGF-B are overtly normal and fertile, but their hearts are reduced in size, suggesting that VEGF-B may have a role in coronary vascularization and growth²². Little is known about the normal physiological role of VEGF-D³.

VEGF must be well regulated

Compared to its more recently discovered relatives, much more is known about VEGF. It is now quite clear that VEGF is such a potent and critical vascular regulator that its dosage must be exquisitely regulated in spatial, temporal and quantitative manner to avoid vascular disaster. Disruption of both VEGF alleles in mice mimicks

Figure 2 Schematic representation of the roles of VEGF, Ang1, Ang2 and ephrin-B2 during vessel formation. The processes include vasculogenesis (stage A), angiogenic remodelling (B), stabilization and maturation (C), destabilization (D), regression (E) and sprouting (F), as described in detail in the text. An attempt is made to assign the indicated vascular growth factors to the various processes, and to indicate their expression patterns. Although not noted in the figure, expression of ephrin-B2 marks arterial vessels from the earliest developmental times.



knockout of VEGFR-2, resulting in almost complete absence of a vasculature^{23,24}. Disruption of even a single VEGF allele in mice leads to embryonic lethality due to severe vascular abnormalities, providing perhaps the only example of embryonic lethality due to a simple half-dosage effect^{23,24}. Even more subtle alterations in VEGF expression during embryonic development result in profound abnormalities, leading to embryonic or early post-natal death^{25,26}. VEGF continues to be critical during early post-natal growth and development, as evidenced by post-natal VEGF inactivation using Cre-loxP-mediated VEGF gene deletion, or by administration of a soluble VEGF receptor that effectively blocks VEGF action²⁷. Although VEGF inactivation is lethal during the first few post-natal weeks, VEGF inactivation in older animals is much less traumatic, seemingly affecting only those structures that continue to undergo vascular remodelling, such as bone growth plates or ovarian corpus lutei^{27–29}. Thus, VEGF does not seem to have a continuous maintenance function for much of the adult vasculature.

The most elegant demonstration of the need for exquisite VEGF regulation involves retinal vascularization, which occurs post-natally in rodents. Angiogenic sprouting into the initially avascular and hypoxic rodent retina depends upon its VEGF expression^{30–32}. Any perturbation of normal VEGF expression patterns destroys retinal vascularization patterns, with dire results for retinal function; subsequent restoration of VEGF expression does not correct the problem, but rather exacerbates it. A simple way to perturb VEGF expression involves exposing post-natal rodents to a brief period of hyperoxia^{31,33,34}, which transiently suppresses retinal VEGF, resulting in cessation of vessel growth and even causing vascular regression^{31,33,34}. When the rodents are returned to normoxia, the now undervascularized retina becomes hypoxic, causing an abnormal burst of VEGF, which promotes robust new angiogenesis, but of haemorrhagic and

leaky vessels growing in totally abnormal patterns that wreak havoc upon the retina. This model reflects the ability of oxygen therapy in premature infants to cause retinopathy of prematurity, and shows the need for precise regulation of VEGF. Similarly, diabetic retinopathy initiates with damage and loss of healthy vessels, followed by retinal hypoxia and resulting VEGF induction, once again leading to an abnormal angiogenic response with leaky and haemorrhagic vessels^{35,36}. These findings show that inappropriate induction of VEGF, in the absence of the entire angiogenic programme, leads to formation of immature and leaky vessels that cause disease. These findings also show that tissue hypoxia cannot necessarily induce a useful angiogenic response.

Consistent with the above findings concerning the devastating consequences of unregulated VEGF expression, several studies have delivered excess VEGF to adult tissues — to adult muscle using retrovirally engineered myoblasts³⁷, to skin using transgenic or adenoviral delivery^{38–41}, or to whole animals using acute adenoviral delivery⁴² — and found that leaky and haemorrhagic vessels were formed, often associated with an inflammatory response, resulting in pronounced tissue swelling and oedema.

The angiopoietins and their Tie receptors

Despite its requisite role in vascular formation, VEGF must work in concert with other factors. The angiopoietins (Fig. 1b) seem to be some of VEGF's most important partners (Fig. 2). The angiopoietins were discovered as ligands for the Ties, a family of receptor tyrosine kinases that are as selectively expressed within the vascular endothelium (despite expression in some other cells, such as in the haemopoietic lineage) as are the VEGF receptors^{43–47}. There are now four definitive members of the angiopoietin family, although Ang3 and Ang4 may represent widely diverged counterparts of the same gene locus in mouse

and man^{12,48,49}. All of the known angiopoietins bind primarily to Tie2, and it is unclear whether there are independent ligands for the second Tie receptor, Tie1, or — as currently seems more likely — whether the known angiopoietins can in some way or under some circumstances also engage Tie1, perhaps as a second component in a heteromerized complex. The rest of this review will deal only with Ang1 and Ang2, since little more can be said at this time about Ang3 and Ang4.

Ang1 stabilizes vessel walls

The most important insights into the normal roles of Ang1 and its Tie2 receptor came from the analysis of mice engineered to lack these gene products^{11,50,51}. Unlike mouse embryos lacking VEGF or VEGFR-2, embryos lacking Ang1 or Tie2 develop a rather normal primary vasculature. However, this vasculature fails to undergo normal further remodelling. The most prominent defects are in the heart, with problems in the associations between the endocardium and underlying myocardium as well as in trabeculae formation, and also in the remodelling of many vascular beds into large and small vessels. In these vascular beds, as in the heart, ultrastructural analysis indicates that endothelial cells fail to associate appropriately with underlying support cells, which are the cells that provide the Ang1 protein that acts on endothelial Tie2 receptors¹¹. This finding led to the suggestion that Ang1 does not supply an instructive signal that actually directs specific vascular remodelling events, but rather has more of a permissive role by optimizing the manner in which endothelial cells integrate with supporting cells, thus allowing them to receive other critical signals from their environment¹¹.

Transgenic overexpression of Ang1 in skin results in pronounced hypervascularization^{40,52}. Although there are modest increases in vessel number, the most marked increase is in vessel size. In contrast, VEGF in similar models primarily increases vessel number^{38–40}. These findings indicate that Ang1 might promote circumferential as opposed to sprout growth. Combining transgenic Ang1 and VEGF leads to unprecedented hypervascularity resulting from increases in both vessel size and number⁴⁰. The vascular patterns induced by the combination are still obviously abnormal morphologically, suggesting that much must be learned about exploiting even this growth factor combination in therapeutic settings so as to grow normal vessels.

In addition to their effects on vascular morphology, transgenic overexpression of Ang1 and VEGF had distinct effects on vascular function and integrity. As had been expected, VEGF led to immature, leaky and haemorrhagic vessels^{38–40}. On the other hand, Ang1 led to vessels that were actually resistant to leak, whether the leak was induced by VEGF or inflammatory agents⁴⁰. This resistance seems related to the ability of Ang1 to maximize interactions between endothelial cells and their surrounding support cells and matrix, as the Ang1 vessels were resistant to treatments that normally created holes in the endothelial cell barrier⁴⁰. These findings indicated that Ang1 might counter the effect of VEGF on permeability, raising multiple therapeutic possibilities⁴⁰. There are numerous disease processes — ranging from diabetic retinopathy to inflammation to brain oedema following ischaemic stroke — in which vessels become damaged and leaky, and an agent that could repair the damage and prevent the leak could have enormous therapeutic benefit. Supporting the clinical potential of Ang1, acute adenoviral administration of Ang1 to adult animals showed that Ang1 can indeed protect the adult vasculature from vascular leak, without inducing immediate changes in vascular morphology⁴².

Ang2: agonist and antagonist?

Ang2 was cloned based on its homology to Ang1, and displayed similarly high affinity for Tie2, but — depending on the cell examined — Ang2 could either activate or antagonize Tie2 (ref. 12). Transgenic overexpression of Ang2 in the embryonic endothelium resulted in embryonic death due to defects resembling those of Ang1 or Tie2 knockouts, demonstrating that Ang2 could act as a Tie2 antagonist *in*

vivo, at least under some circumstances¹². This possibility became even more intriguing when Ang2 expression profiles were examined. In adult animals, Ang2 was induced in the endothelium of vessels undergoing active remodelling, such as sprouting or regressing vessels in the ovary^{12,53}, or in tumours^{13,14,54,55} (as will be discussed in detail below). These findings, together with the possibility that Ang2 could act as a Tie2 antagonist, led to the hypothesis that Ang2 might provide a key de-stabilizing signal involved in initiating angiogenic remodelling^{12–14,55}. That is, based on previous evidence that Ang1 engagement of the Tie2 receptor was constitutive in the adult vasculature and indeed necessary to maintain its quiescence (Fig. 2, stage C), it was proposed that autocrine induction of Ang2 in endothelium blocked this constitutive stabilizing influence of paracrine Ang1, allowing the endothelial cells to revert to a more plastic and destabilized state reminiscent of developing vessels (Fig. 2, stage D). Such destabilized vessels could then be prone to two fates. On the one hand, these destabilized vessels would be prone to regression in the absence of associated growth factors, as also occurs with primitive vessels during development (Fig. 2, stage E). On the other hand, they would be more sensitive to angiogenic changes induced by simultaneously available angiogenic factors such as VEGF, essentially recapitulating an early embryonic situation in which VEGF acts prior to the involvement of Ang1 (Fig. 2, stage F).

This model of Ang2 as a destabilizing signal that reverts vessels to a more plastic and tenuous state, initially developed based on observations in the remodelling ovary¹², is consistent with more recent data in tumours (see below) as well as emerging data from knockout mice lacking Ang2. One of the best characterized settings of post-natal vascular regression and remodelling in mice involves the eye, in which regression of the hyaloid vasculature encasing the lens is coupled to angiogenic sprouting that leads to vascularization of the initially avascular retina, as described above. Neither regression of the hyaloid vasculature nor vascularization of the retina occur in mice lacking Ang2 (S.J.W., R. Tzekova, Q. Wong, N.W.G., C. Suri & G.D.Y., unpublished results). These data show that Ang2 is required for some post-natal vascular remodelling events, and support the notion that Ang2 provides a key role in destabilizing the vasculature in a manner that is necessary for its subsequent remodelling. However, other defects in the Ang2-knockout mice suggest that it may in some cases also have an agonistic role. That is, it is highly expressed in the developing aortic wall, which does not develop properly in mice lacking Ang2. Similarly, lymphatic development is perturbed in these mice.

The ephrins

The Eph receptor tyrosine kinases comprise the largest known family of growth factor receptors (Fig. 1c), and use the similarly numerous ephrins as their ligands^{7,56}. The ephrins are unlike ligands for other receptor tyrosine kinases in that they must be tethered to the membrane to activate their Eph receptors^{7,57}. Although initially characterized in the nervous system^{7,56}, recent knockout studies have suggested key roles for ephrin-B2 and its EphB4 receptor during vascular development^{58–60}. Mouse embryos lacking ephrin-B2 and EphB4 suffer fatal defects in early angiogenic remodelling that are somewhat reminiscent of those seen in mice lacking Ang1 or Tie2^{58–60}. Moreover, ephrin-B2 and EphB4 display remarkably reciprocal distribution patterns during vascular development, with ephrin-B2 marking the endothelium of primordial arterial vessels while EphB4 marks the endothelium of primordial venous vessels^{58–60}. These distributions suggested that ephrin-B2 and EphB4 are involved in establishing arterial versus venous identity, perhaps in fusing arterial and venous vessels at their junctions, and that defects in these processes might account for the early lethality observed in mouse embryos lacking these proteins^{58–60} (Fig. 2, stage A).

Ephrin-B2 continues to selectively mark arteries during later embryonic development as well as in the adult, although this expression extends progressively from the arterial endothelium to the

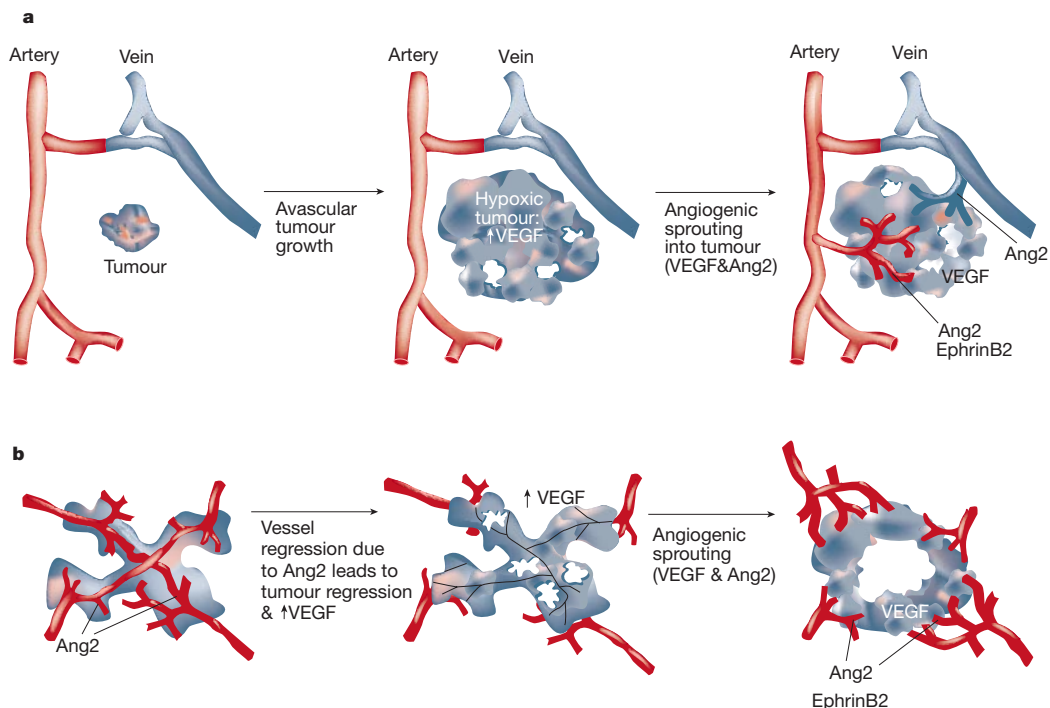


Figure 3 Models of tumour angiogenesis. **a**, Model of avascular tumour initiation contrasted with **b**, tumour initiation involving host vessel co-option. An attempt is made to assign the indicated vascular growth factors to roles in the various indicated steps in tumour development, and to indicate their expression patterns.

surrounding arterial smooth muscle and to pericytes (N.W.G. and G.D.Y., unpublished results; D. Shin and D. J. Anderson, unpublished results). Thus, ephrin-B2 is apparently not only required during the earliest stages of arterial/venous determination, but may continue to be important during the development of arteries, perhaps by regulating interactions between endothelial and smooth muscle cells involved in the formation of arterial muscular walls (Fig. 2, stage B). In adult settings of angiogenesis, as in tumours or in the female reproductive system, the endothelium of new vessels strongly re-expresses ephrin-B2 (N.W.G. and G.D.Y., unpublished results; D. Shin and D. J. Anderson, unpublished results) (Fig. 3a,b). The finding that angiogenic sprouting in the adult and in tumours involves re-expression of the ephrin-B2 arterial marker challenges existing dogma that such sprouting primarily involves venous or uncommitted vessels, and also suggests that ephrin-B2 may be important in these angiogenic settings.

VEGF and Ang2 in tumour angiogenesis

Much has been made of the notion that tumours and metastases initiate as small avascular masses, which only subsequently induce the angiogenic ingrowth that is required to allow further growth of the early tumour^{61–63} (Fig. 3a). It is clear that many natural tumours initially arise in this manner, particularly primary epithelial tumours that are initially separated from underlying vessels by a basement membrane that must be broken before tumour cells can access the vasculature. In addition, many artificial model systems forcibly create initially avascular tumours by placing tumour cells in a space that is normally devoid of vessels — such as the subcutaneous space, the cornea pocket or the vitreous or the tumour window — thus requiring angiogenesis to get vessels to the tumour.

Despite all the attention directed towards avascular tumour growth, recent findings^{14,55} have refocused attention on previous observations^{64–66} that many tumours, and metastases in particular, do not initiate in an avascular manner (Fig. 3b). Rather, tumour cells can initially home in on and grow by co-opting existing host vessels,

and thus start off as well-vascularized small tumours^{13,14} (Fig. 3b, left). In response to co-option, the host vessels mount a defence — sensing inappropriate co-option, they regress, choking off the tumour and resulting in a secondarily avascular and hypoxic tumour (Fig. 3b, middle). However, successful tumours seem to overcome host vessel regression by inducing robust new angiogenesis (Fig. 3b, right). Ang2 and VEGF inductions correlate remarkably well with the above processes^{13,14,55}. That is, soon after tumour co-option, host vessels start expressing high autocrine levels of Ang2; thus Ang2 is one of the earliest tumour markers described, and one of the most general because it marks co-opted vessels and not the tumour cells themselves (Fig. 3b, left). Consistent with the possibility that autocrine Ang2 expression can destabilize vessels (Fig. 2, stage D), the co-opted vessels begin to die by an apoptotic process shortly after expressing Ang2 (Fig. 3b, middle). As vessels die, the tumour becomes secondarily avascular and hypoxic, resulting in marked induction of tumour-derived VEGF (Fig. 3b, middle). These high levels of VEGF correlate with cessation of regression of the destabilized co-opted vessels, and onset of robust new angiogenesis sprouting from these vessels, allowing for tumour survival and further growth (Fig. 3b, right). Thus, in such settings, endothelial Ang2 expression seems to correlate with vessel destabilization, apparently leading to vessel regression in the absence of tumour-derived VEGF, or robust new angiogenesis following induction of tumour-derived VEGF (stage D in Fig. 2, and Fig. 3b). The possibility that tumour vessel Tie2 receptors are blocked continuously by Ang2 and thus have an imbalance towards VEGF may well explain long-standing observations that tumour vessels fail to mature, exhibit poor associations between endothelial cells and their supporting cells, and are characterized by their leaky and haemorrhagic state.

One practical prediction, which applies whether tumour growth initiates avascularly or through co-option, is that anti-VEGF therapy should ultimately blunt tumour growth. Early studies using an anti-VEGF antibody provided the first support for this notion⁶⁷. This

has subsequently been confirmed in many laboratories using numerous approaches ranging from antibodies that bind and block VEGF, to those that bind and block VEGFR-2, to small molecules that block the activity of the VEGF-2 kinase domain, to genetic ablation of VEGF in tumour cells⁶⁸. Thus, blockade of VEGF represents the best validated and most compelling anti-angiogenesis approach described so far.

Perspectives and therapeutic possibilities

There are many critical growth factors involved in the physiological regulation of blood vessel formation, and the actions of these molecular players must be very carefully orchestrated in terms of time, space and dose so as to form a functioning vascular network. The complexity of the process makes ongoing therapeutic efforts aimed at growing new vascular networks to treat ischaemic disease, using random delivery of single agents, appear somewhat naive with the potential to cause more harm (by forming malfunctioning vessels prone to leak and haemorrhage) than good. In their defence, these efforts were initiated years ago when much less was understood about the process of vascular formation. Recent failures of large, well-controlled clinical trials for cardiac ischaemia using delivery of single agents (either VEGF or FGF)^{69,70} raises the question of why these trials failed despite claims of success in animal studies and earlier, smaller (and uncontrolled) human trials. As recently discussed⁶⁸, this may be due to the failure of animal models to correctly model the human disease, as well as the need for blind approaches in both animal and human studies to overcome investigator bias when measuring subjective endpoints, together with the requirement for placebo controls in settings where there is a marked placebo effect in subjective patient reports of their own condition.

Although the complexities of vascular formation create significant challenges for those trying to grow vessels for therapeutic use, these same complexities may work in favour of therapeutic approaches aimed at blocking vessel growth. That is, blockade of many different molecular players may all result in the blunting of vessel formation. There is no doubt that VEGF is the best-validated target for anti-angiogenesis therapies, based on overwhelming genetic, mechanistic and animal efficacy data. Despite the attention devoted to a number of other putative angiogenic antagonists for use in cancer (for example, endostatin, angiostatin and antithrombin)^{71–73}, most of these antagonists have yet to be characterized from a mechanistic and genetic point of view. Thus, they lack defined mechanisms of action, and cannot be placed within existing models of molecular angiogenesis using genetic approaches. Also troubling is that these agents seem to work whether they are delivered as properly folded proteins or as denatured aggregates⁷².

Recent efforts also indicate as yet unimagined applications for vascular growth factors. For example, the possibility that Ang1 may help prevent or repair damaged and leaky vessels offers therapeutic hope for an assortment of unmet clinical needs, such as in diabetic retinopathy, acute macular degeneration, ischaemia/reperfusion injury (which can occur after strokes and in acute respiratory distress syndrome), or in inflammatory settings^{40,42}. The continued discovery and characterization of the molecular factors that regulate vessel formation will lead to additional unexpected therapeutic opportunities, as well as to the refinement of current therapeutic approaches aimed at growing or blocking vessel formation. □

1. Ferrara, N. Vascular endothelial growth factor: molecular and biological aspects. *Curr. Top. Microbiol. Immunol.* **237**, 1–30 (1999).
2. Dvorak, H. F., Nagy, J. A., Feng, D., Brown, L. F. & Dvorak, A. M. Vascular permeability factor/vascular endothelial growth factor and the significance of microvascular hyperpermeability in angiogenesis. *Curr. Top. Microbiol. Immunol.* **237**, 97–132 (1999).
3. Eriksson, U. & Alitalo, K. Structure, expression and receptor-binding properties of novel vascular endothelial growth factors. *Curr. Top. Microbiol. Immunol.* **237**, 41–57 (1999).
4. Risau, W. Mechanisms of angiogenesis. *Nature* **386**, 671–674 (1997).
5. Kenyon, B. M. *et al.* A model of angiogenesis in the mouse cornea. *Invest. Ophthalmol. Vis. Sci.* **37**, 1625–1632 (1996).
6. Ribatti, D., Vacca, A., Roncali, L. & Dammacco, F. The chick embryo chorioallantoic membrane as a

- model for in vivo research on angiogenesis. *Int. J. Dev. Biol.* **40**, 1189–1197 (1996).
7. Gale, N. W. & Yancopoulos, G. D. Growth factors acting via endothelial cell-specific receptor tyrosine kinases: VEGFs, angiopoietins, and ephrins in vascular development. *Genes Dev.* **13**, 1055–1066 (1999).
8. Carmeliet, P. Mechanisms of angiogenesis and arteriogenesis. *Nature Med.* **6**, 389–395 (2000).
9. Hellstrom, M., Kaln, M., Lindahl, P., Abramsson, A. & Betsholtz, C. Role of PDGF-B and PDGFR-beta in recruitment of vascular smooth muscle cells and pericytes during embryonic blood vessel formation in the mouse. *Development* **126**, 3047–3055 (1999).
10. Hirschi, K. K., Rohovsky, S. A., Beck, L. H., Smith, S. R. & D'Amore, P. A. Endothelial cells modulate the proliferation of mural cell precursors via platelet-derived growth factor-BB and heterotypic cell contact. *Circ. Res.* **84**, 298–305 (1999).
11. Suri, C. *et al.* Requisite role of Angiopoietin-1, a ligand for the Tie2 receptor, during embryonic angiogenesis. *Cell* **87**, 1171–1180 (1996).
12. Maisonpierre, P. C. *et al.* Angiopoietin-2, a natural antagonist for Tie2 that disrupts in vivo angiogenesis. *Science* **277**, 55–60 (1997).
13. Holash, J., Wiegand, S. J. & Yancopoulos, G. D. New model of tumor angiogenesis: dynamic balance between vessel regression and growth mediated by angiopoietins and VEGF. *Oncogene* **18**, 5356–5362 (1999).
14. Holash, J. *et al.* Vessel cooption, regression, and growth in tumors mediated by angiopoietins and VEGF. *Science* **284**, 1994–1998 (1999).
15. Soker, S., Takashima, S., Miao, H., Neufeld, G. & Klagsbrun, M. Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor. *Cell* **92**, 735–745 (1998).
16. Shalaby, F. *et al.* Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature* **376**, 62–66 (1995).
17. Fong, G. H., Rossant, J., Gertsenstein, M. & Breitman, M. L. Role of the Flt-1 receptor tyrosine kinase in regulating assembly of vascular endothelium. *Nature* **376**, 66–70 (1995).
18. Hiratsuka, S., Minowa, O., Kuno, J., Noda, T. & Shibuya, M. Flt-1 lacking the tyrosine kinase domain is sufficient for normal development and angiogenesis in mice. *Proc. Natl Acad. Sci. USA* **95**, 9349–9354 (1998).
19. Taipale, J. *et al.* Vascular endothelial growth factor receptor-3. *Curr. Top. Microbiol. Immunol.* **237**, 85–96 (1999).
20. Persico, M. G., Vincenti, V. & DiPalma, T. Structure, expression and receptor-binding properties of placenta growth factor (PlGF). *Curr. Top. Microbiol. Immunol.* **237**, 31–40 (1999).
21. Olofsson, B., Jeltsch, M., Eriksson, U. & Alitalo, K. Current biology of VEGF-B and VEGF-C. *Curr. Opin. Biotechnol.* **10**, 528–535 (1999).
22. Bellomo, D. *et al.* Mice lacking the vascular endothelial growth factor-B gene (Vegfb) have smaller hearts, dysfunctional coronary vasculature, and impaired recovery from cardiac ischemia. *Circ. Res.* **86**, E29–E35 (2000).
23. Carmeliet, P. *et al.* Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* **380**, 435–439 (1996).
24. Ferrara, N. *et al.* Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* **380**, 439–442 (1996).
25. Carmeliet, P. *et al.* Impaired myocardial angiogenesis and ischemic cardiomyopathy in mice lacking the vascular endothelial growth factor isoforms VEGF164 and VEGF188. *Nature Med.* **5**, 495–502 (1999).
26. Miquelot, L., Langille, B. L. & Nagy, A. Embryonic development is disrupted by modest increases in vascular endothelial growth factor gene expression. *Development* **127**, 3941–3946 (2000).
27. Gerber, H. P. *et al.* VEGF is required for growth and survival in neonatal mice. *Development* **126**, 1149–1159 (1999).
28. Ferrara, N. *et al.* Vascular endothelial growth factor is essential for corpus luteum angiogenesis. *Nature Med.* **4**, 336–340 (1998).
29. Gerber, H. P. *et al.* VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation. *Nature Med.* **5**, 623–628 (1999).
30. Stone, J. *et al.* Development of retinal vasculature is mediated by hypoxia-induced vascular endothelial growth factor (VEGF) expression by neuroglia. *J. Neurosci.* **15**, 4738–4747 (1995).
31. Alon, T. *et al.* Vascular endothelial growth factor acts as a survival factor for newly formed retinal vessels and has implications for retinopathy of prematurity. *Nature Med.* **1**, 1024–1028 (1995).
32. Benjamin, L. E., Hemo, I. & Keshet, E. A plasticity window for blood vessel remodelling is defined by pericyte coverage of the preformed endothelial network and is regulated by PDGF-B and VEGF. *Development* **125**, 1591–1598 (1998).
33. Stone, J. *et al.* Roles of vascular endothelial growth factor and astrocyte degeneration in the genesis of retinopathy of prematurity. *Invest. Ophthalmol. Vis. Sci.* **37**, 290–299 (1996).
34. Pierce, E. A., Foley, E. D. & Smith, L. E. Regulation of vascular endothelial growth factor by oxygen in a model of retinopathy of prematurity. *Arch. Ophthalmol.* **114**, 1219–1228 (1996). [Published erratum appears in *Arch. Ophthalmol.* **115**, 427 (1997).]
35. Aiello, L. P. *et al.* Vascular endothelial growth factor in ocular fluid of patients with diabetic retinopathy and other retinal disorders. *N. Engl. J. Med.* **331**, 1480–1487 (1994).
36. Adamis, A. P. *et al.* Increased vascular endothelial growth factor levels in the vitreous of eyes with proliferative diabetic retinopathy. *Am. J. Ophthalmol.* **118**, 445–450 (1994).
37. Springer, M. L., Chen, A. S., Kraft, P. E., Bednarski, M. & Blau, H. M. VEGF gene delivery to muscle: potential role for vasculogenesis in adults. *Mol. Cell* **2**, 549–558 (1998).
38. Detmar, M. *et al.* Increased microvascular density and enhanced leukocyte rolling and adhesion in the skin of VEGF transgenic mice. *J. Invest. Dermatol.* **111**, 1–6 (1998).
39. Larcher, F., Murillas, R., Bolontrade, M., Conti, C. J. & Jorcano, J. L. VEGF/VPF overexpression in skin of transgenic mice induces angiogenesis, vascular hyperpermeability and accelerated tumor development. *Oncogene* **17**, 303–311 (1998).
40. Thurston, G. *et al.* Leakage-resistant blood vessels in mice transgenically overexpressing angiopoietin-1. *Science* **286**, 2511–2514 (1999).
41. Pettersson, A. *et al.* Heterogeneity of the angiogenic response induced in different normal adult tissues by vascular permeability factor/vascular endothelial growth factor. *Lab. Invest.* **80**, 99–115 (2000).
42. Thurston, G. *et al.* Angiopoietin-1 protects the adult vasculature against plasma leakage. *Nature Med.* **6**, 460–463 (2000).
43. Korhonen, J. *et al.* Enhanced expression of the tie receptor tyrosine kinase in endothelial cells during

- neovascularization. *Blood* **80**, 2548–2555 (1992).
44. Maisonnier, P. C., Goldfarb, M., Yancopoulos, G. D. & Gao, G. Distinct rat genes with related profiles of expression define a TIE receptor tyrosine kinase family. *Oncogene* **8**, 1631–1637. (1993).
45. Sato, T. N., Qin, Y., Kozak, C. A. & Audus, K. L. *tie-1* and *tie-2* define another class of putative receptor tyrosine kinase genes expressed in early embryonic vascular system. *Proc. Natl Acad. Sci. USA* **90**, 9355–9358 (1993).
46. Dumont, D. J., Gradwohl, G. J., Fong, G.-H., Auerbach, R. & Breitman, M. L. The endothelial-specific receptor tyrosine kinase, *tek*, is a member of a new subfamily of receptors. *Oncogene* **8**, 1293–1301 (1993).
47. Iwama, A. *et al.* Molecular cloning and characterization of mouse *Tie* and *Tek* receptor tyrosine kinase genes and their expression in hematopoietic stem cells. *Biochem. Biophys. Res. Commun.* **195**, 301–309 (1993).
48. Davis, S. *et al.* Isolation of angiopoietin-1, a ligand for the TIE2 receptor, by secretion-trap expression cloning. *Cell* **87**, 1161–1169 (1996).
49. Valenzuela, D. *et al.* Angiopoietins 3 and 4: diverging gene counterparts in mouse and man. *Proc. Natl Acad. Sci. USA* **96**, 1904–1909 (1999).
50. Dumont, D. J. *et al.* Dominant-negative and targeted null mutations in the endothelial receptor tyrosine kinase, *tek*, reveal a critical role in vasculogenesis of the embryo. *Genes Dev.* **8**, 1897–1909 (1994).
51. Sato, T. N. *et al.* Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation. *Nature* **376**, 70–74 (1995).
52. Suri, C. *et al.* Angiopoietin-1 promotes increased vascularization in vivo. *Science* **282**, 468–471 (1998).
53. Goede, V., Schmidt, T., Kimmina, S., Kozian, D. & Augustin, H. G. Analysis of blood vessel maturation processes during cyclic ovarian angiogenesis. *Lab. Invest.* **78**, 1385–1394 (1998).
54. Stratmann, A., Risau, W. & Plate, K. H. Cell type-specific expression of Angiopoietin-1 and Angiopoietin-2 suggests a role in glioblastoma angiogenesis. *Am. J. Pathol.* **153**, 1459–1466 (1998).
55. Zagzag, D. *et al.* In Situ expression of angiopoietins in astrocytomas identifies angiopoietin-2 as an early marker of tumor angiogenesis. *Exp. Neurol.* **159**, 391–400 (1999).
56. Flanagan, J. G. & Vanderhaeghen, P. The ephrins and Eph receptors in neural development. *Annu. Rev. Neurosci.* **21**, 309–345 (1998).
57. Davis, S. *et al.* Ligands for the EPH-related receptor tyrosine kinases that require membrane attachment or clustering for activity. *Science* **266**, 816–819 (1994).
58. Wang, H. U., Chen, Z. & Anderson, D. J. Molecular distinction and angiogenic interaction between embryonic arteries and veins revealed by ephrin-B2 and its receptor Eph-B4. *Cell* **93**, 741–753 (1998).
59. Adams, R. H. *et al.* Roles for ephrinB ligands and EphB receptors in cardiovascular development: demarcation of arterial/venous domains, vascular morphogenesis, and sprouting angiogenesis. *Genes Dev.* **13**, 295–306 (1999).
60. Gerety, S. S., Wang, H. U., Chen, Z. F. & Anderson, D. J. Symmetrical mutant phenotypes of the receptor EphB4 and its specific transmembrane ligand ephrin-B2 in cardiovascular development. *Mol. Cell* **4**, 403–414 (1999).
61. Folkman, J. Tumor angiogenesis: therapeutic implications. *N. Engl. J. Med.* **285**, 1182–1186 (1971).
62. Folkman, J. What is the evidence that tumors are angiogenesis dependent? *J. Natl. Cancer Inst.* **82**, 4–6 (1990).
63. Hanahan, D. & Folkman, J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* **86**, 353–364 (1996).
64. Wesseling, P., van der Laak, J. A., de Leeuw, H., Ruiter, D. J. & Burger, P. C. Quantitative immunohistological analysis of the microvasculature in untreated human glioblastoma multiforme. *J. Neurosurg.* **81**, 902–909 (1994).
65. Holmgren, L., O'Reilly, M. S. & Folkman, J. Dormancy of micrometastases: balanced proliferation and apoptosis in the presence of angiogenesis suppression. *Nature Med.* **1**, 149–153 (1995).
66. Pezzella, F. *et al.* Non-small-cell lung carcinoma tumor growth without morphological evidence of neo-angiogenesis. *Am. J. Pathol.* **151**, 1417–1423 (1997).
67. Kim, K. J. *et al.* Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth in vivo. *Nature* **362**, 841–844 (1993).
68. Ferrara, N. & Alitalo, K. Clinical applications of angiogenic growth factors and their inhibitors. *Nature Med.* **5**, 1359–1364 (1999).
69. Henry, T. D. *et al.* Results of intracoronary recombinant human vascular endothelial growth factor (RhVEGF) administration trial. *J. Am. Coll. Cardiol.* **31**, 65A (1998).
70. Chiron Corporation. Chiron announces preliminary findings from phase II clinical trial of FGF-2. <<http://www.prnewswire.com/micro/CHIR>> Company Press Release 12 March 2000.
71. O'Reilly, M. S. *et al.* Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a Lewis Lung Carcinoma. *Cell* **79**, 315–328 (1994).
72. O'Reilly, M. S. *et al.* Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell* **88**, 277–285 (1997).
73. O'Reilly, M. S., Pirie-Shepherd, S., Lane, W. S. & Folkman, J. Antiangiogenic activity of the cleaved conformation of the serpin antithrombin. *Science* **285**, 1926–1928 (1999).



Angiogenesis in cancer and other diseases

Peter Carmeliet* & Rakesh K. Jain†

*The Center for Transgene Technology and Gene Therapy, Flanders Interuniversity Institute for Biotechnology, KU Leuven, Leuven, B-3000, Belgium (e-mail: peter.carmeliet@med.kuleuven.ac.be)

†Steele Laboratory for Tumor Biology, Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA (e-mail: jain@steele.mgh.harvard.edu)

Pathological angiogenesis is a hallmark of cancer and various ischaemic and inflammatory diseases. Concentrated efforts in this area of research are leading to the discovery of a growing number of pro- and anti-angiogenic molecules, some of which are already in clinical trials. The complex interactions among these molecules and how they affect vascular structure and function in different environments are now beginning to be elucidated. This integrated understanding is leading to the development of a number of exciting and bold approaches to treat cancer and other diseases. But owing to several unanswered questions, caution is needed.

Mammalian cells require oxygen and nutrients for their survival and are therefore located within 100 to 200 μm of blood vessels — the diffusion limit for oxygen. For multicellular organisms to grow beyond this size, they must recruit new blood vessels by vasculogenesis and angiogenesis (Box 1 Fig.). This process is regulated by a balance between pro- and anti-angiogenic molecules, and is derailed in various diseases, especially cancer (Table 1). Without blood vessels, tumours can not grow beyond a critical size or metastasize to another organ. Similarly, without an efficient blood supply we may not be able to deliver anti-cancer drugs to all regions of a tumour in effective quantities. With advances in molecular genetics and the availability of molecular probes and imaging technologies, we are now obtaining insight into physiological and pathological angiogenesis. In this review, we discuss how normal and abnormal blood vessels form, how they function, what key molecules (genes) are involved and how they are used for therapy, why caution is warranted and what key questions remain unanswered.

Tumour angiogenesis

The angiogenic switch

The observation that angiogenesis occurs around tumours was made nearly 100 years ago^{1–3}. The hypothesis that tumours produce a diffusible ‘angiogenic’ substance was put forward in 1968^{4,5}. In 1971, Folkman proposed that tumour growth and metastasis are angiogenesis-dependent, and hence, blocking angiogenesis could be a strategy to arrest tumour growth⁶. This possibility stimulated an intensive search for pro- and anti-angiogenic molecules (Table 2). In 1976, Gullino showed that cells in pre-cancerous tissue acquire angiogenic capacity on their way to becoming cancerous⁷. He proposed that this concept be used to design strategies to prevent cancer⁷, a hypothesis later confirmed by genetic approaches⁸.

It is now widely accepted that the ‘angiogenic switch’ is ‘off’ when the effect of pro-angiogenic molecules is balanced by that of anti-angiogenic molecules, and is ‘on’ when the net balance is tipped in favour of angiogenesis^{8,9}.

Various signals that trigger this switch have been discovered (see Table 2). These include metabolic stress (for example, low $p\text{O}_2$, low pH or hypoglycaemia), mechanical stress (for example, pressure generated by proliferating cells), immune/inflammatory response (for example, immune/inflammatory cells that have infiltrated the tissue), and genetic mutations (for example, activation of oncogenes or deletion of tumour-suppressor genes that control production of angiogenesis regulators)^{10,11}. How the interplay between environmental and genetic mechanisms influences tumour angiogenesis and growth is a complex and largely unresolved matter. Pro- and anti-angiogenic molecules can emanate from cancer cells, endothelial cells, stromal cells, blood and the extracellular matrix¹². Their relative contribution is likely to change with tumour type and tumour site. It is also likely to change with tumour growth, regression and relapse. The challenge now is to establish a unified framework incorporating quantitative data on the magnitude and temporal sequence of the generation of these molecules¹³. This should help develop effective therapeutic strategies.

Formation of tumour vessels

Vessels in an embryo are assembled from endothelial precursors (vasculogenesis). Subsequently, this primitive network expands by sprouting (angiogenesis) or intussusception, in which interstitial tissue columns are inserted into the lumen of pre-existing vessels and partition the vessel lumen¹⁴. Tumour vessels develop by sprouting or intussusception from pre-existing vessels. Circulating endothelial precursors, shed from the vessel wall or mobilized from the bone marrow, can also contribute to tumour angiogenesis^{15,16}. Tumour cells can also grow around an existing vessel to form a perivascular cuff (see review in this issue by Yancopoulos *et al.*, pages 242–248). We now know there are various molecular players involved in these different mechanisms of vascular growth¹⁷ (Box 1). Among these, members of the vascular endothelial growth factor (VEGF) and angiopoietin (Ang) family have a predominant role (see review by Yancopoulos *et al.*). The angiogenic activity of VEGF is tightly regulated by gene dosage^{18,19}. Several molecules, including a number of angiogenesis inhibitors, seem to be involved mainly in

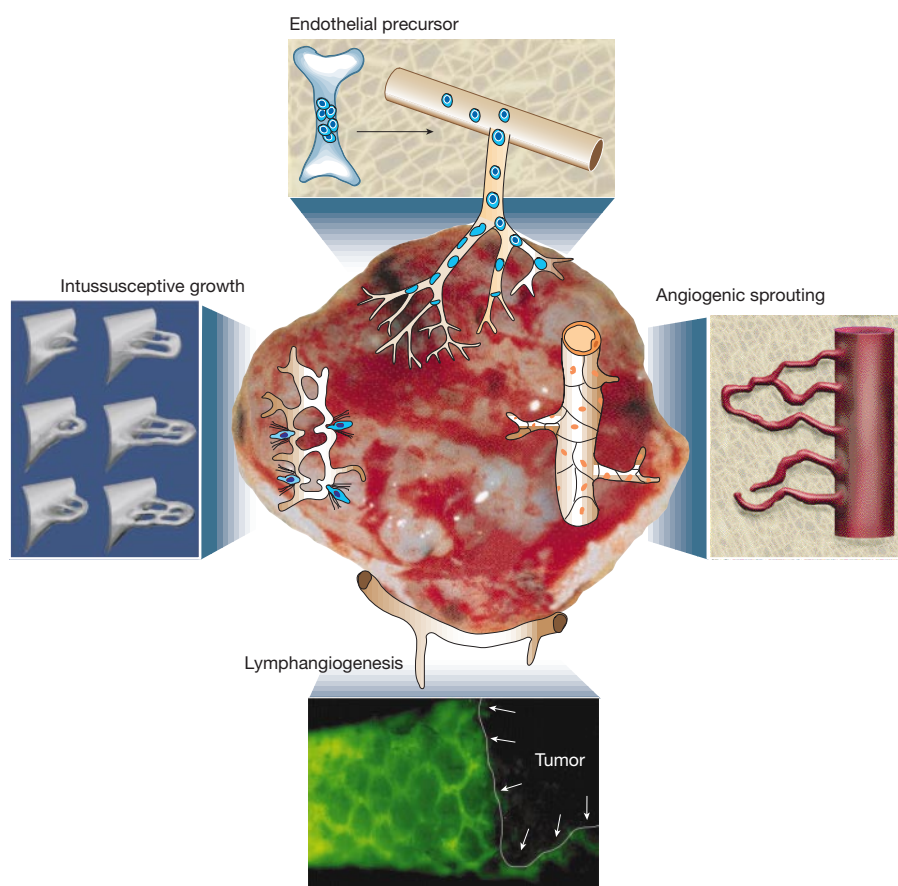
Box 1

Formation of tumour vessels: molecular and cellular mechanisms

Tumour vessels can grow by sprouting, intussusception or by incorporation of bone marrow-derived endothelial precursors. In addition, tumour cells can co-opt existing vessels. Several molecules have been implicated in these processes (Table 2; see review in this issue by Yancopoulos *et al.*, pages 242–248, and ref. 17). During sprouting angiogenesis, vessels initially dilate and become leaky in response to VEGF. Ang1 and the junctional molecules VE-cadherin and platelet-endothelial cell-adhesion molecule (PECAM) tighten vessels and their action needs to be overcome during angiogenesis. Ang2 and proteinases mediate dissolution of the existing basement membrane and the interstitial matrix. Numerous molecules stimulate endothelial proliferation, migration and assembly, including VEGF, Ang1 and bFGF (Table 2). Cell-matrix receptors such as the $\alpha_v\beta_3$ and α_5 integrins mediate cell spreading and migration. Maturation of nascent vessels involves formation of a new basement membrane and investment of new vessels with pericytes and smooth muscle cells. PDGF-BB recruits smooth muscle cells, whereas signalling by TGF- β 1 and Ang1/Tie2 stabilizes the interaction between endothelial and smooth muscle cells. Proteinase inhibitors (for example, PAI-1) prevent degradation of the provisional extracellular matrix around nascent vessels. Maintenance of new vessels depends on the survival of endothelial cells. In a normal adult, quiescent endothelial cells can survive for several years. VEGF (through an interaction with VE-cadherin⁷⁰) and Ang1 are vital survival factors. In contrast, most angiogenesis inhibitors cause endothelial apoptosis. By binding VEGF, soluble VEGF receptors (for example, VEGFR-1, neuropilin-1) reduce the angiogenic activity of VEGF. Molecules that initially induce angiogenesis are subsequently (proteolytically) processed to angiogenesis inhibitors, thereby providing a negative feedback. Most angiogenesis inhibitors suppress tumour angiogenesis; their role in normal vascular growth remains largely unknown.

VEGF, bFGF, granulocyte macrophage-colony stimulating factor (GM-CSF), IGF-1 and angiopoietins have been implicated in the mobilization of endothelial precursors, whereas angiopoietins are important in vessel co-option. Several molecules are only involved in tumour angiogenesis (among them, $\alpha_v\beta_3$, PAI-1, NO, cyclo-oxygenase-2 (COX-2), thrombospondin-2 (TSP-2) and a large list of angiogenesis inhibitors (Table 2). The mechanism of action of some of these regulators is poorly understood. For instance, although proteinases might be expected to stimulate tumour angiogenesis by ‘clearing the path’ for migrating endothelial cells, the proteinase inhibitor PAI-1 is a poor prognostic factor. Indeed, PAI-1 is required to prevent uncontrolled plasmin proteolysis, as this causes widespread matrix dissolution and prevents endothelial assembly^{71,72}.

Box 1 Figure Cellular mechanisms of tumour (lymph) angiogenesis. Tumour vessels grow by various mechanisms: (1) the host vascular network expands by budding of endothelial sprouts or formation of bridges (angiogenesis); (2) tumour vessels remodel and expand by the insertion of interstitial tissue columns into the lumen of pre-existing vessels (intussusception); and (3) endothelial cell precursors (angioblasts) home from the bone marrow or peripheral blood into tumours and contribute to the endothelial lining of tumour vessels (vasculogenesis). Lymphatic vessels around tumours drain the interstitial fluid and provide a gateway for metastasizing tumour cells. (Adapted from ref. 38.)



tumour angiogenesis (Table 2). The temporal and spatial expression of these regulators is not as well coordinated in tumours as in physiological angiogenesis, and their mechanism of action is poorly understood (Box 1 Fig.). In addition, tumour vessels lack protective mechanisms that normal vessels acquire during growth. For example, they may lack functional perivascular cells, which are

needed to protect vessels against changes in oxygen or hormonal balance, provide them necessary vasoactive control to accommodate metabolic needs, and induce vascular quiescence²⁰. Finally, the vessel wall is not always formed by a homogenous layer of endothelial cells²¹. Instead, it may be lined with only cancer cells or a mosaic of cancer and endothelial cells. For example, 15% of vessels

Table 1 Angiogenesis in neoplasms and other diseases

Organ	Processes characterized by abnormal angiogenesis or vascular malfunction*	Organ	Processes characterized by abnormal angiogenesis or vascular malfunction*
Blood vessels	†Atherosclerosis, haemangioma, haemangioendothelioma §Vascular malformations	Bone, joints	†Rheumatoid arthritis, synovitis, bone and cartilage destruction, osteomyelitis, pannus growth, osteophyte formation, cancer ‡Aseptic necrosis, impaired healing of fractures
Skin	†Warts, pyogenic granulomas, hair growth, Kaposi's sarcoma, scar keloids, allergic oedema, neoplasms §Psoriasis (skin vessels enlarge and become tortuous) ‡Decubitus or stasis ulcers, gastrointestinal ulcers	Liver, kidney, lung, ear and other epithelia	†Inflammatory and infectious processes (hepatitis, pneumonia, glomerulonephritis), asthma, nasal polyps, transplantation, liver regeneration, cancer §Pulmonary hypertension, diabetes ‡Pulmonary and systemic hypertension (vascular pruning)
Uterus, ovary, placenta	†Dysfunctional uterine bleeding (contraception), follicular cysts, ovarian hyperstimulation, endometriosis, neoplasms §Pre-eclampsia ‡Placental insufficiency	Brain, nerves, eye	†Retinopathy of prematurity, diabetic retinopathy, choroidal and other intraocular disorders, leukomalacia, cancer ‡Stroke, vascular dementia, Alzheimer's disease, CADASIL
Peritoneum, pleura	Respiratory distress, ascites, peritoneal sclerosis (dialysis patients), adhesion formation (abdominal surgery), metastatic spreading	Endocrine organs	†Thyroiditis, thyroid enlargement, pancreas transplantation ‡Thyroid pseudocyst
Heart, skeletal muscle	†Work overload ‡Ischaemic heart and limb disease	Lymph vessels	†Tumour metastasis, lymphoproliferative disorders ‡Lymphoedema
Adipose tissue	†Obesity	Haematopoiesis	†AIDS (Kaposi), haematologic malignancies

*List of selected examples.

†Increased vascularization; ‡insufficient vascularization; §abnormal remodelling; ||increased vascularization and/or permeability; see text for abbreviations.

in xenografted and spontaneous human colon carcinomas are mosaic in nature²² (Fig. 1a,b). It is an open question whether these vessels result from cancer cells invading the vessel lumen, from cancer cells mimicking endothelial cells ('vasculogenic mimicry'), from co-opted vessels or from the apoptosis of endothelial cells which exposes underlying cancer cells. Regardless of the mechanism involved, the presence of cancer cells in tumour vessels has significant implications for metastasis and for the design of anti-angiogenic therapy.

Structure and function of tumour vessels

Chaotic architecture and blood flow. Tumour vessels are structurally and functionally abnormal. In contrast to normal vessels, tumour vasculature is highly disorganized, vessels are tortuous and dilated, with uneven diameter, excessive branching and shunts (Fig. 1c,d). This may be due to an imbalance of angiogenic regulators, such as VEGF and angiopoietins. Consequently, tumour blood flow is chaotic and variable²³ and leads to hypoxic and acidic regions in tumours²⁴ (Box 2 and Fig. 2). These conditions lower therapeutic effectiveness, modulate the production of angiogenic stimulators and inhibitors, and select for cancer cells that are more malignant and metastatic. In addition, hypoxia may select for clonal expansion of cells that have lost their apoptotic response to hypoxia²⁵. Finally, although smooth muscle α -actin positive cells surround some tumour vessels, they do not function as normal contractile cells, making the pharmacological alteration of tumour blood flow a challenge^{21,26}.

High vascular permeability. In terms of their ultrastructure, tumour vessels are also abnormal: their walls have numerous 'openings' (endothelial fenestrae, vesicles and transcellular holes), widened interendothelial junctions, and a discontinuous or absent basement membrane (Fig. 1c,d). In addition, the endothelial cells are abnormal in shape, growing on top of each other and projecting into the lumen. These defects make tumour vessels leaky²⁷⁻²⁹. However, there is tremendous heterogeneity in leakiness over space and time and in response to treatment³⁰. Vascular permeability and angiogenesis depend on the type of tumour and the host organ where the tumour is growing^{27,31}, in part because each organ has different stromal cells which produce different pro- and anti-angiogenic molecules^{12,32} (Box 1). Low-permeability tumours may overexpress Ang1 and/or underexpress VEGF or its homologue, placental growth factor (PlGF). Conversely, those with high permeability may lack Ang1 or overexpress its antagonist Ang2 (ref. 33). To overcome this heterogeneity, which is a major challenge for cancer treatment, further studies are needed to discern how angiogenic molecules cooperate.

Non-uniform surface markers. Cytokines and angiogenic molecules secreted by cancer and immune cells can modulate the expression of cellular adhesion molecules and other surface markers on the tumour endothelium. For example, VEGF and tumour-necrosis factor- α (TNF- α) upregulate, whereas basic fibroblast growth factor (bFGF) and transforming growth factor- β 1 (TGF- β 1) downregulate adhesion molecules³⁴. Chaotic blood supply coupled with non-uniform expression of adhesion molecules may explain why leukocyte-endothelial interaction is low in tumours and why activated lymphocytes adhere non-uniformly to tumour vessels. It is possible that tumour vessels express surface proteins that are absent or barely detectable in mature vessels^{35,36}. *In vivo* selection of phage display libraries has recently yielded peptides (for example, amino-acid sequences RGD and NGR) that preferentially recognize vessels in subcutaneous tumours in mice³⁷. These peptides can be used to target therapeutic agents to tumours. The challenge now is to discern how specific these 'vascular zip codes' are, as targeting drugs to the tumour vasculature has the potential to change the paradigm for cancer treatment.

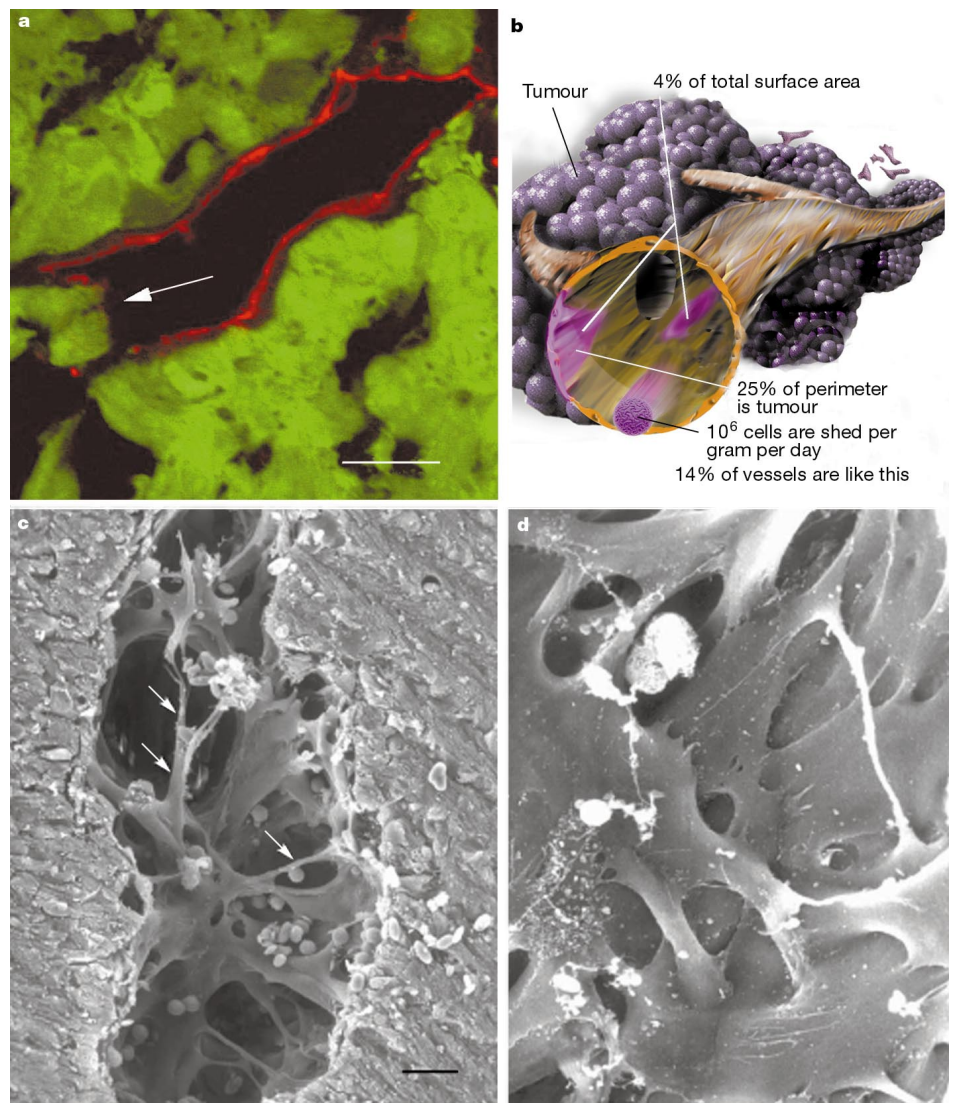
Lack of functional lymphatics

Lymphangiography of a transplanted fibrosarcoma shows functional lymphatics are absent within the tumour and enlarged at the periphery (Box 1 Fig.)³⁸. One explanation may be that neoplastic cells grown in a confined space generate mechanical stress, which may compress the newly formed lymphatic channels inside the tumour³⁹, whereas at the periphery, excess VEGF-C causes lymphatics to enlarge⁴⁰. These enlarged lymphatics may collect interstitial fluid and metastatic cancer cells 'oozing' from the tumour surface, and thus facilitate lymphatic metastasis. Absence of functional lymphatics within tumours may contribute to interstitial hypertension and interfere with the delivery of therapeutic agents⁴¹. Further studies are needed to block lymphatic metastasis and overcome the pressure barrier to the delivery of molecular medicine in tumours.

Tumour dormancy

Human tumours can remain dormant for years owing to a balance between cell proliferation and apoptosis. As a result of their longer half-life, the systemic concentration of angiogenic inhibitors may exceed that of stimulators and inhibit growth of metastases at distal sites. This hypothesis formed the basis of the discovery of angiostatin, endostatin, vasculostatin and other endogenous inhibitors of angiogenesis^{42,43} (Table 2). Most of these inhibitors were discovered by growing primary tumours subcutaneously, a site that is not natural for most human tumours³². Indeed, Gohongi *et al.*⁴⁴ showed that

Figure 1 Chaotic and mosaic vessels in tumours. **a**, Cancer cell in the lining of a tumour vessel, referred to as a mosaic vessel. Cellular components of the vascular wall in a human colon carcinoma xenograft: cancer cells (green fluorescence, green fluorescent protein (GFP)), endothelial cells (red fluorescence, CD31/CD105 antibody detected with cyanine 5) and lectin fluorescence to mark perfused vessels. The width of the endothelial gap exposing cancer cells to the vessel lumen is about 20 μm . (Adapted from ref. 22). **b**, Quantification of mosaic vessels. In colon carcinoma ~15% of tumour vessels are mosaic in nature, and cancer cells occupy ~4% of the total vascular surface area. If each of these cells intravasate in 2 days, the tumour will shed about 10^6 cells per day per gram of tumour (Adapted from ref. 22). **c**, **d**, Scanning electron microscopy of the luminal surface of a blood vessel in a murine mammary tumour showing various abnormalities. Bar length represents 15 μm (from ref. 28). **c**, The abnormal endothelial cells that partition the lumen (arrowheads); **d**, multiple intercellular openings (arrows) of the order of 1–5 μm .



production of angiogenesis inhibitors, similar to angiogenesis stimulators, is dependent on the site of the primary tumour. This is not surprising given that these factors can be produced directly by host or tumour cells or as a result of cleavage of extracellular proteins by enzymes produced by these cells. The production of these inhibitors may also change during the course of therapy. For example, radiation has been shown to increase the production of various angiogenic molecules including endostatin in tumours⁴⁵. These findings have important implications for combined anti-angiogenic and cytotoxic therapies.

Haematological malignancies and haemangiomas

Compared with the single, straight microvessels in normal bone marrow, complex branching of microvessels has been observed in leukaemic bone marrow⁴⁶. Malignant haematopoietic cells produce and also respond to various angiogenic factors such as VEGF and angiopoietins⁴⁷. Vascular tumours (haemangiomas) in children are quite common and incapacitating, but we know little about their aetiology⁴⁸. VEGFs stimulate, whereas interferons inhibit their growth⁶. Haemangiomas may arise from genetic alterations or viral infections. For instance, products of human herpesvirus-8 and human immunodeficiency virus type-1 (HIV-1) have been implicated in the pathogenesis of Kaposi's sarcoma, found in ~30% of AIDS patients^{49,50}. HIV-1 Tat protein activates VEGFR-2, binds endothelial $\alpha_5\beta_1$ and $\alpha_v\beta_3$ integrins and retrieves bFGF from the

extracellular matrix⁴⁹. It is thus possible that anti-angiogenic drugs will expand the arsenal for the treatment of leukaemia and AIDS.

Angiogenesis in non-neoplastic diseases

In a normal adult, most vasculature is quiescent, with only 0.01% of endothelial cells undergoing division. Excessive or insufficient vascular growth contributes to numerous non-neoplastic disorders, and the list is growing rapidly (Table 1). In other diseases, vessels do not grow, but rather abnormally remodel. While tumour angiogenesis is primarily an 'endothelial disorder', vascular growth and remodelling affect both endothelial and smooth muscle cells. Inflammation and hypoxia contribute to angiogenesis in non-neoplastic diseases.

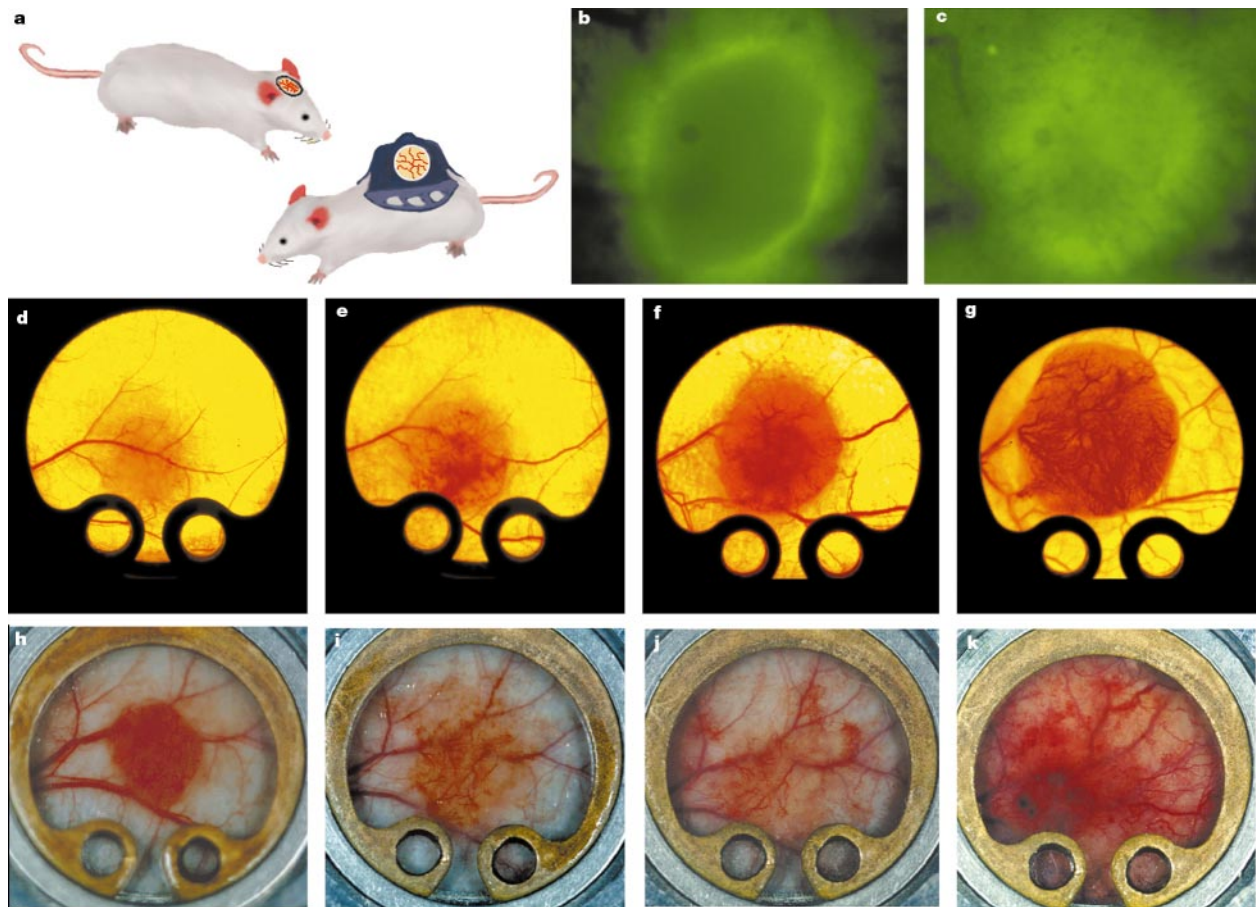
Hypoxia-driven pathological angiogenesis and vascular remodelling

Hypoxia is a strong stimulus for angiogenesis in numerous disorders. Cells in tumours, wounds or atherosclerotic plaques become hypoxic when too distant from nearby vessels. Abnormal deposition of extracellular matrix or vascular congestion impairs delivery of oxygen and causes hypoxia in diabetes⁵¹, Alzheimer's disease and asthma. The supply of oxygen can also become limited by vascular pruning during hypertension or upon exposure of premature babies to high oxygen. Recent discoveries have shown that hypoxia activates hypoxia-inducible transcription factors (HIFs), which function as master switches to induce expression of several angiogenic factors including VEGF, nitric oxide synthase (NOS), platelet-derived growth factor

Box 2

Intravital microscopy to visualize gene expression and vessel function

Intravital microscopy allows non-invasive imaging of gene expression and vessel formation as well as of the architecture and function of the resulting vessels. Usually these parameters are measured with approaches that require tissue removal. These invasive methods are unable to capture the time course of events in the same tissue, and owing to tissue heterogeneity, a relatively large number of samples need to be analysed to reach conclusions. In intravital microscopy, the tissue of interest is visualized through a surgically implanted, chronic glass window (for example, dorsal skin or cranium; panels **a**, **b** in figure below) or by surgical manipulation/exteriorization of the tissue (for example, liver or mesentery)⁷³. Gene expression or cell lineage is monitored with a live fluorescent reporter (for example, GFP or RFP)¹² (panels **c**, **d**). The images are captured continuously on a highly sensitive video camera attached to a light or fluorescent microscope and stored on videotape. Computer-assisted analysis of these images can yield parameters such as vessel morphology (diameter, length and tortuosity), haemodynamics (blood flow rate), pH, pO_2 , vascular permeability, leukocyte adhesion, and microscopic distribution of fluorescently labelled molecules, particles and cells. By growing a tumour in different organs, the effect of host–tumour interaction on gene expression and physiological function can be examined⁷⁴. Finally, these parameters can be monitored during tumour growth (panels **e–h**) as well as during tumour regression and relapse (panels **i–l**).



Box 2 Figure Intravital microscopy. **a**, Cranial window. **b**, Dorsal window. **c**, **d**, VEGF promoter activity, as monitored by GFP intensity, on day 7 and 14 in murine mammary carcinoma growing in the dorsal window (adapted from ref. 12). **e–h**, Angiogenesis and tumour growth in a 5-, 10-, 15- and 20-day-old human colon adenocarcinoma in the dorsal window (adapted from ref. 75). **i–l**, Angiogenesis during tumour regression and relapse following hormone ablation therapy of a hormone (androgen/testosterone)-dependent murine mammary carcinoma (adapted from ref. 30).

(PDGF), Ang2 and others (Fig. 2)⁵². Hypoxia-driven angiogenesis salvages ischaemic myocardium and prolongs survival of stroke patients. However, it can cause blindness in premature newborns and in diabetic patients⁵³, and haemorrhagic rupture of atherosclerotic plaques. Apart from stimulating angiogenesis, hypoxia can also cause vascular remodelling. In chronic obstructive lung disease, hypoxia causes irreversible loss of vessels and thickening of the vascular muscular coat, with resultant life-threatening pulmonary

hypertension⁵⁴. This vascular remodelling has been ascribed to an imbalance between vasodilators (nitric oxide) and vasoconstrictors (endothelin-1).

Inflammation-induced angiogenesis and vascular remodelling

Prolonged and excessive angiogenesis is a hallmark of inflammatory disorders in many organs (Table 1). Monocytes, macrophages, platelets, mast cells and other leukocytes release a myriad of angiogenic factors including VEGF, Ang1, bFGF, TGF- β 1, PDGF, TNF- α ,

Table 2 Angiogenesis activators and inhibitors			
Activators	Function	Inhibitors	Function
VEGF family members†‡	Stimulate angio/vasculogenesis, permeability, leukocyte adhesion	VEGFR-1; soluble VEGFR-1; soluble NRP-1	Sink for VEGF, VEGF-B, PlGF
VEGFR†, NRP-1	Integrate angiogenic and survival signals	Ang2†‡	Antagonist of Ang1
Ang1 and Tie2†‡	Stabilize vessels, inhibit permeability	TSP-1,-2	Inhibit endothelial migration, growth, adhesion and survival
PDGF-BB and receptors	Recruit smooth muscle cells	Angiostatin and related plasminogen kringle	Suppress tumour angiogenesis
TGF-β1*, endoglin, TGF-β receptors	Stimulate extracellular matrix production	Endostatin (collagen XVIII fragment)	Inhibit endothelial survival and migration
FGF, HGF, MCP-1	Stimulate angio/arteriogenesis	Vasostatin; calreticulin	Inhibit endothelial growth
Integrins αβ ₃ , αβ ₂ , αβ ₁	Receptors for matrix macromolecules and proteinases	Platelet factor-4	Inhibits binding of bFGF and VEGF
VE-cadherin; PECAM (CD31)	Endothelial junctional molecules	TIMPs; MMP inhibitors; PEX	Suppress pathological angiogenesis
Ephrins‡	Regulate arterial/venous specification	Meth-1; Meth-2	Inhibitors containing MMP, TSP and disintegrin domains
Plasminogen activators, MMPs	Remodel matrix, release and activate growth factors	IFN-α, -β, -γ; IP-10, IL-4, IL-12, IL-18	Inhibit endothelial migration; downregulate bFGF
PAI-1	Stabilize nascent vessels	Prothrombin kringle-2; antithrombin III fragment	Suppress endothelial growth
NOS; COX-2	Stimulate angiogenesis and vasodilation	Prolactin (M _r 16K)	Inhibits bFGF/VEGF
AC133	Regulate angioblast differentiation	VEGI	Modulate cell growth
Chemokines*	Pleiotropic role in angiogenesis	Fragment of SPARC	Inhibit endothelial binding and activity of VEGF
Id1/Id3	Determine endothelial plasticity	Osteopontin fragment	Interfere with integrin signalling
		Maspin	Protease inhibitor
		Canstatin, proliferin-related protein, restin	Mechanisms unknown

List of selected examples; further information and references are available at <http://steele.mgh.harvard.edu>. Abbreviations: VEGFR, VEGF receptors; MMP, matrix metalloproteinase; Id1/3, inhibitors of differentiation 1/3; NRP-1, neuropilin-1; TIMPs, tissue inhibitors of MMP; PEX, proteolytic fragment of MMP2; IP-10, inducible protein-10; VEGI, member of TNF family; SPARC, inhibits endothelial binding and activity of VEGF; other abbreviations defined in the text.
*Opposite effect in some contexts.
†Also present in or affecting non-endothelial cells.
‡See review in this issue by Yancopoulos *et al.*, pages 242–248.

hepatocyte growth factor (HGF), insulin-like growth factor-1 (IGF-1), monocyte chemoattractant protein-1 (MCP-1), among many others^{55,56}. Some of these factors attract wound cells, which in turn release additional angiogenic factors^{57,58}. Blood cells also contain proteinases that degrade anatomical barriers for migrating vascular cells⁵⁹, and activate or liberate some of these growth factors from the extracellular matrix^{58,60}. Haematopoietic cells are also involved in the negative control of angiogenesis. For instance, they release inhibitors such as platelet-factor 4 and thrombospondins, and cause proteolytic conversion of plasminogen to angiostatin, and collagen XVIII to

endostatin^{42,43}. The challenge now is to reveal how the temporal sequence and degree of infiltration of these different blood-borne cells in wounds or tumours determine the angiogenic switch. Vasodilation and increased permeability during inflammation — in the absence of angiogenesis — can be predominant pathogenic mechanisms. For instance, oedema contributes to infarct expansion after stroke and may cause life-threatening intracranial hypertension in cancer patients. Furthermore, extravasation of plasma proteins favours metastatic spread of occult tumours, and airway congestion may cause fatal asthmatic attacks.

Figure 2 Role of hypoxia in tumour angiogenesis. Because of the irregular pattern and organization of the tumour vasculature, some cells in tumours are located more than 100 μm (the diffusion limit for oxygen) away from blood vessels and become hypoxic (red-to-blue gradient indicates progressive hypoxia). Tumour cells survive fluctuations in oxygen tensions, in part because clones are selected in hypoxic tumours that switch to a proangiogenic phenotype. HIFs increase transcription of several angiogenic genes (for example, genes encoding VEGF, PDGF-BB and NOS). HIFs also affect cellular survival/apoptosis pathways. Inset: relationship between the distance of tumour cells from nearby vessels and their degree of hypoxia (blue symbols) and acidosis (red symbols)²⁴.

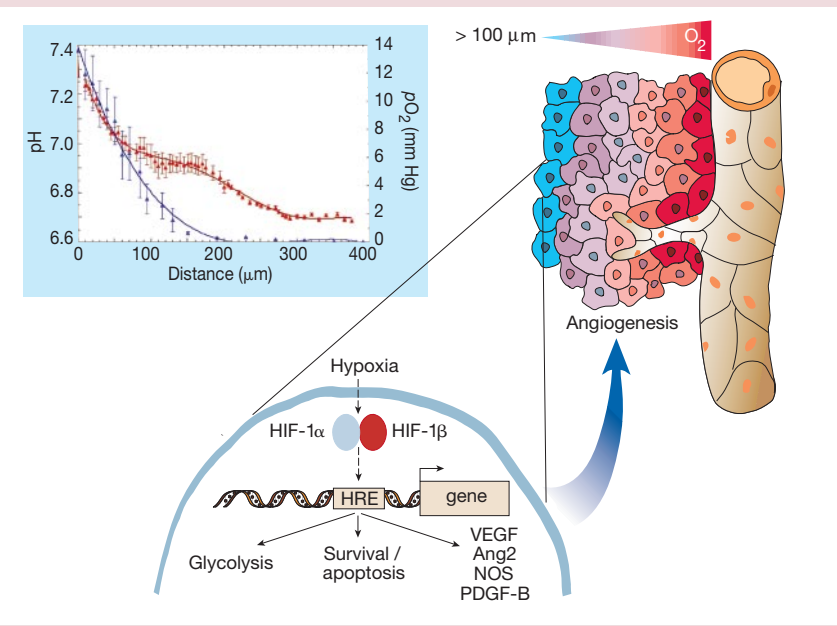


Table 3 Angiogenesis inhibitors in clinical trials for cancer

Drug	Mechanism
Phase I	
EMD121974	Small molecule integrin antagonist
Combretastatin A-4 prodrug	Apoptosis in proliferating endothelium
PTK787/ZK2284	Blocks VEGF-receptor signalling
Endostatin	Induces endothelial cell apoptosis <i>in vivo</i>
BMS-275291	Synthetic MMP inhibitor
SU6668	Blocks VEGF-, FGF- and PDGF-receptor signalling
Phase II	
CAI	Inhibitor of calcium influx
Squalamine	Inhibits Na ⁺ /H ⁺ exchanger
COL-3	Synthetic MMP inhibitor: tetracycline derivative
CGS-27023A	Synthetic MMP inhibitor
TNP-470	Fumagillin analogue; inhibits endothelial proliferation
Vitaxin	Antibody to integrin on endothelial surface
IL-12	Induces interferon- γ and IP-10
Anti-VEGF Ab	Monoclonal antibody to VEGF
Phase III	
SU5416	Blocks VEGF receptor signalling
Thalidomide	Unknown
Marimastat	Synthetic MMP inhibitor
AG3340	Synthetic MMP inhibitor
Neovastat	Natural MMP inhibitor
Interferon- α	Inhibition of bFGF and VEGF production
IM862	Unknown mechanism

From NCI Database <www.cancertrials.nci.nih.gov> (updated 12 April 2000).

Inflammatory cells have also been implicated in other vascular process, for example, the growth of pre-existing collateral arterioles after occlusion of a supply artery in the myocardium and peripheral limbs^{57,61}. This process has been termed 'adaptive arteriogenesis' to distinguish it from true 'angiogenesis' (capillary growth). Monocytes are recruited by elevated levels of MCP-1. These cells infiltrate and proteolytically remodel the vessel wall⁵⁷. Local production of bFGF, PDGF-B and TGF- β 1 stimulates growth of endothelial and smooth muscle cells. Adaptive arteriogenesis finally results in functional and structurally normal arteries, which ameliorate the detrimental effects of vessel obstruction⁶¹. These vessels may be superior to newly formed capillaries (formed by angiogenesis), because they are able to sustain proper circulation and adapt to changes in physiological demands of blood supply.

Vascular malformations

Vascular anomalies are localized lesions of dysmorphic blood or lymphatic vessels⁴⁸. Some of these lesions have an inherited predisposition, often resulting from mutations in genes involved in the cross-talk between endothelial and smooth muscle cells. For example, abnormal Tie2 signalling has been linked to reduced smooth muscle cell recruitment in venous malformations. Impaired stabilization of vessels due to defects in TGF- β 1 signalling predisposes to arteriovenous malformations in patients with hereditary haemorrhagic telangiectasia. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL) results from abnormal signalling by Notch3, a gene implicated in interactions between endothelial and smooth muscle cells. Mutations in VEGF receptor-3, a candidate lymphangiogenic player, cause congenital lymphoedema¹⁹.

Angiogenesis and obesity

Angiogenesis may contribute to excess accumulation of body fat in obese individuals. Indeed, preadipocytes migrate to sites of neovascularization and adipose tissue is highly angiogenic⁶². VEGF, bFGF (induced by insulin) and leptin (a central mediator in obesity) have been identified as mediators of angiogenesis in adipose tissue⁶³. The possibility of using anti-angiogenic therapy to treat obese patients is an important area of further investigation.

Therapeutic perspectives: promises and problems

Therapeutic angiogenesis

Various angiogenic approaches to treat ischaemic diseases are already in clinical trials^{19,64}. Most interventions involve the delivery of VEGF or bFGF to the ischaemic tissue to stimulate growth of new vessels. Vessels formed by VEGF are leaky and tortuous. It may be possible to control leakiness by combining VEGF with Ang1, but the diameter of resulting vessels is not uniform. Whether this abnormal vascular morphology can lead to impaired microcirculation is not known³³. Furthermore, it is not known whether increased systemic levels of angiogenic cytokines during the course of these therapies will alter the expression of adhesion molecules in systemic circulation, trigger dormant tumours, and/or accelerate atherosclerosis.

Another outstanding question is whether a single angiogenic factor will be able to stimulate 'functional and sustainable' angiogenesis or if a combination of angiogenic molecules will be required. For example, genetic studies showed that the VEGF₁₂₀ isoform alone is able to initiate, but not complete, the angiogenic programme⁶⁵. Because hypoxia-inducible factors initiate an entire angiogenic response, they have been considered for angiogenic (gene) therapy in ischaemic conditions⁶⁶. However, caution is warranted as these hypoxia-inducible factors could also control cell death⁶⁷. These problems notwithstanding, therapeutic angiogenic approaches offer new hope for illnesses that were previously considered intractable.

Anti-angiogenic therapies

Based on successful preclinical data, several anti-angiogenic agents alone or in combination with conventional therapies are now in clinical trials⁶ (Table 3; www.cancertrials.nci.nih.gov). These trials are based on strategies that (1) interfere with angiogenic ligands, their receptors or downstream signalling; (2) upregulate or deliver endogenous inhibitors, or (3) directly target tumour vasculature. These approaches offer new hope for the successful treatment of cancer. However, there are a number of potential problems that warrant caution in clinical trials on humans.

First, most preclinical studies are carried out in tumours that are grown subcutaneously, which is not a common site for human tumours. Because the host organ can alter the biology of the tumour, further studies are needed with spontaneous or orthotopically grown tumours to more accurately predict the response of human tumours. A second problem with most preclinical studies is that they use tumour regression, not eradication as an end-point. Because relapse can occur from a very small number of surviving cancer cells, recurrence must be accounted for in treatment study designs. Often, the response (growth delay) is measured for too short a time and treatment is started before the tumour is established. Thus, studies are needed to help predict the long-term response of clinically established tumours. Third, some animal studies use therapy that is toxic only to rapidly proliferating cells (for example, low-dose chemotherapy given over a longer period). In these studies, the proliferation rates of cancer cells and endothelial cells are very high (that is, with a doubling time of the order of days). Thus, preclinical studies are needed in slow-growing spontaneous or orthotopically grown tumours, which are more typical of tumours found in humans⁶⁸.

As tumours grow, they begin to produce a wider array of angiogenic molecules. Therefore, if only one molecule (for example, VEGF) is blocked, tumours may switch to another molecule (for example, bFGF or interleukin (IL)-8). Thus, we may require a cocktail of antibodies/inhibitors. Similarly, in approaches that target tumour vasculature, it is generally assumed that most endothelial cells in tumours express the same vascular marker ('zip code'). Given the microvascular heterogeneity in tumours, studies are needed to test the validity of this assumption. If the tumour vascular marker is also present in normal tissues, strategies are needed to prevent injury to these tissues. Furthermore, we need to determine whether cells lining tumour vessels are always non-transformed and do not develop drug resistance to long-term anti-angiogenic treatment. For

example, deletion of the transcription factor HIF-1 α decreases VEGF, angiogenesis and pO₂ and yet results in some tumour types that grow more rapidly. Will similar mutations make tumours resistant to anti-angiogenic therapy?

The long-term side effects of many anti-angiogenic therapies on normal tissues and physiological angiogenesis are not known. For example, mice expressing a mutant VEGF gene, in which the hypoxia-response element was deleted, suffer adult-onset motor neuron degeneration like patients with amyotrophic lateral sclerosis (B. Oosthuysen *et al.*, unpublished results). Along with desired response in tumours, long-term interference with VEGF signalling may cause tumour-dependent normal tissue toxicity. Although vessel count has been shown to be a successful prognostic factor in many human tumours, it does not predict vascular function. In fact, vessels become more efficient (that is, develop a normal phenotype) during the early phase of some anti-angiogenic therapies. Furthermore, tumours do not 'shrink' during various anti-angiogenic therapies. Thus, new imaging methods are needed to monitor vascular function and, consequently, therapeutic response in patients. Finally, the angiogenic response to various stimuli depends on the individual genetic constitution of experimental animals⁶⁹. Therefore, we may have to tailor treatments to individual pharmacogenetic profiles.

Many of these problems can be addressed with further carefully planned animal studies. The solution to others will become apparent from the ongoing clinical trials. In the meantime, anti-angiogenic therapy is the most promising approach to cancer treatment. It can potentially overcome two major problems associated with other therapies — the problem of poor delivery⁴¹ and the problem of drug resistance^{6,10}. Most importantly, it is hypothesis-driven. □

- Goldman, E. The growth of malignant disease in man and the lower animals with special reference to the vascular system. *Lancet* **2**, 1236–1240 (1907).
- Ide, A. G., Baker, N. H. & Warren, S. L. Vascularization of the Brown-Pearce rabbit epithelioma transplant as seen in the transparent ear chamber. *Am. J. Radiol.* **42**, 891–899 (1939).
- Algire, G. H. & Chalkley, H. W. Vascular reactions of normal and malignant tissues in vivo. I. Vascular reactions of mice to wounds and to normal and neoplastic transplants. *J. Natl Cancer Inst. USA* **6**, 73–85 (1945).
- Greenblatt, M. & Shubik, P. Tumor angiogenesis: transfilter diffusion studies in the hamster by the transparent chamber technique. *J. Natl Cancer Inst.* **41**, 111–124 (1968).
- Ehrmann, R. L. & Knoth, M. Choriocarcinoma: transfilter stimulation of vasoproliferation in the hamster cheek pouch studied by light and electron microscopy. *J. Natl Cancer Inst.* **41**, 1329–1341 (1968).
- Folkman, J. in *Cancer Medicine* (eds Holland, J. F. *et al.*) 132–152 (Decker, Ontario, Canada, 2000).
- Gullino, P. M. Angiogenesis and oncogenesis. *J. Natl Cancer Inst.* **61**, 639–643 (1978).
- Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
- Bouck, N., Stellmach, V. & Hsu, S. C. How tumors become angiogenic. *Adv. Cancer Res.* **69**, 135–174 (1996).
- Kerbel, R. S. Tumor angiogenesis: past, present and the near future. *Carcinogenesis* **21**, 505–515 (2000).
- Carmeliet, P. Controlling the cellular brakes. *Nature* **401**, 657–658 (1999).
- Fukumura, D. *et al.* Tumor induction of VEGF promoter activity in stromal cells. *Cell* **94**, 715–725 (1998).
- Ramanujan, S., Koenig, G. C., Padera, T. P., Stoll, B. R. & Jain, R. K. Local imbalance of proangiogenic and antiangiogenic factors: a potential mechanism of focal necrosis and dormancy in tumors. *Cancer Res.* **60**, 1442–1448 (2000).
- Patan, S., Munn, L. L. & Jain, R. K. Intussusceptive microvascular growth in a human colon adenocarcinoma xenograft: a novel mechanism of tumor angiogenesis. *Microvasc. Res.* **51**, 260–272 (1996).
- Asahara, T., Kalka, C. & Isner, J. M. Stem cell therapy and gene transfer for regeneration. *Gene Ther.* **7**, 451–457 (2000).
- Rafii, S. Circulating endothelial precursors: mystery, reality, and promise. *J. Clin. Invest.* **105**, 17–19 (2000).
- Carmeliet, P. Mechanisms of angiogenesis and arteriogenesis. *Nature Med.* **6**, 389–395 (2000).
- Carmeliet, P. *et al.* Abnormal blood vessel development and lethality in embryos lacking a single vascular endothelial growth factor allele. *Nature* **380**, 435–439 (1996).
- Ferrara, N. & Allitalo, K. Clinical applications of angiogenic growth factors and their inhibitors. *Nature Med.* **5**, 1359–1364 (1999).
- Benjamin, L. E., Golijanin, D., Itin, A., Podes, D. & Keshet, E. Selective ablation of immature blood vessels in established human tumors follows vascular endothelial growth factor withdrawal. *J. Clin. Invest.* **103**, 159–165 (1999).
- Jain, R. K. Determinants of tumor blood flow: a review. *Cancer Res.* **48**, 2641–2658 (1988).
- Chang, Y. S. *et al.* Abundance of neoplastic cells in vessel walls of human tumor xenografts. *Proc. Am. Assoc. Cancer Res.* (in the press).
- Baish, J. W. & Jain, R. K. Fractals and cancer. *Cancer Res.* **60**, 3683–3688 (2000).
- Helminger, G., Yuan, F., Dellian, M. & Jain, R. K. Interstitial pH and pO₂ gradients in solid tumors in vivo: high-resolution measurements reveal a lack of correlation. *Nature Med.* **3**, 177–182 (1997).
- Giacca, A. J. Hypoxic stress proteins: survival of the fittest. *Semin. Radiat. Oncol.* **6**, 46–58 (1996).
- Eberhard, A. *et al.* Heterogeneity of angiogenesis and blood vessel maturation in human tumors: implications for antiangiogenic tumor therapies. *Cancer Res.* **60**, 1388–1393 (2000).
- Hobbs, S. K. *et al.* Regulation of transport pathways in tumor vessels: role of tumor type and microenvironment. *Proc. Natl Acad. Sci. USA* **95**, 4607–4612 (1998).
- Hashizume, H. *et al.* Openings between defective endothelial cells explain tumor vessel leakiness. *Am. J. Pathol.* **156**, 1363–1380 (2000).
- Dvorak, H. F., Nagy, J. A., Feng, D., Brown, L. F. & Dvorak, A. M. Vascular permeability factor/vascular endothelial growth factor and the significance of microvascular hyperpermeability in angiogenesis. *Curr. Top. Microbiol. Immunol.* **237**, 97–132 (1999).
- Jain, R. K. *et al.* Endothelial cell death, angiogenesis, and microvascular function after castration in an androgen-dependent tumor: role of vascular endothelial growth factor. *Proc. Natl Acad. Sci. USA* **95**, 10820–10825 (1998).
- Fukumura, D., Yuan, F., Monsky, W. L., Chen, Y. & Jain, R. K. Effect of host microenvironment on the microcirculation of human colon adenocarcinoma. *Am. J. Pathol.* **151**, 679–688 (1997).
- Fidler, I. J. Modulation of the organ microenvironment for treatment of cancer metastasis. *J. Natl Cancer Inst.* **87**, 1588–1592 (1995).
- Jain, R. K. & Munn, L. L. Leaky vessels? Call Angi! *Nature Med.* **6**, 131–132 (2000).
- Jain, R. K. *et al.* Leukocyte-endothelial adhesion and angiogenesis in tumors. *Cancer Metastasis Rev.* **15**, 195–204 (1996).
- Eliceiri, B. P. & Cheresh, D. A. The role of α v integrins during angiogenesis: insights into potential mechanisms of action and clinical development. *J. Clin. Invest.* **103**, 1227–1230 (1999).
- Huang, X. *et al.* Tumor infarction in mice by antibody-directed targeting of tissue factor to tumor vasculature. *Science* **275**, 547–550 (1997).
- Arap, W., Pasqualini, R. & Ruoslahti, E. Cancer treatment by targeted drug delivery to tumor vasculature in a mouse model. *Science* **279**, 377–380 (1998).
- Leu, A. J., Berk, D. A., Lymboussaki, A., Alitalo, K. & Jain, R. K. Absence of functional lymphatics within a murine sarcoma: a molecular and functional evaluation. *Cancer Res.* **60**, 4324–4327 (2000).
- Helminger, G., Netti, P. A., Lichtenfeld, H. C., Melder, R. J. & Jain, R. K. Solid stress inhibits the growth of multicellular tumor spheroids. *Nature Biotechnol.* **15**, 778–783 (1997).
- Jeltsch, M. *et al.* Hyperplasia of lymphatic vessels in VEGF-C transgenic mice. *Science* **276**, 1423–1425 (1997).
- Jain, R. K. Barriers to drug delivery in solid tumors. *Sci. Am.* **271**, 58–65 (1994).
- O'Reilly, M. S. *et al.* Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a Lewis lung carcinoma. *Cell* **79**, 315–328 (1994).
- O'Reilly, M. S. *et al.* Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell* **88**, 277–285 (1997).
- Gohongi, T. *et al.* Tumor-host interactions in the gallbladder suppress distal angiogenesis and tumor growth: involvement of transforming growth factor β 1. *Nature Med.* **5**, 1203–1208 (1999).
- Hartford, A. C., Gohongi, T., Fukumura, D. & Jain, R. K. Irradiation of a primary tumor, unlike surgical removal, enhances angiogenesis suppression at a distal site: potential role of host-tumor interaction. *Cancer Res.* **60**, 2128–2131 (2000).
- Perez Atayde, A. R. *et al.* Spectrum of tumor angiogenesis in the bone marrow of children with acute lymphoblastic leukemia. *Am. J. Pathol.* **150**, 815–821 (1997).
- Bellamy, W. T., Richter, L., Frutiger, Y. & Grogan, T. M. Expression of vascular endothelial growth factor and its receptors in hematopoietic malignancies. *Cancer Res.* **59**, 728–733 (1999).
- Vikkula, M., Boon, L., Mulliken, J. B. & Olsen, B. R. Molecular basis of vascular anomalies. *Trends Cardiovasc. Med.* **8**, 281–292 (1998).
- Albini, A. *et al.* The angiogenesis induced by HIV-1 tat protein is mediated by the Flk-1/KDR receptor on vascular endothelial cells. *Nature Med.* **2**, 1371–1375 (1996).
- Flore, O. *et al.* Transformation of primary human endothelial cells by Kaposi's sarcoma-associated herpesvirus. *Nature* **394**, 588–592 (1998).
- Boulton, A. J. & Malik, R. A. Diabetic neuropathy. *Med. Clin. North Am.* **82**, 909–929 (1998).
- Semenza, G. L. Hypoxia-inducible factor 1: master regulator of O₂ homeostasis. *Curr. Opin. Genet. Dev.* **8**, 588–594 (1998).
- Alon, T. *et al.* Vascular endothelial growth factor acts as a survival factor for newly formed retinal vessels and has implications for retinopathy of prematurity. *Nature Med.* **1**, 1024–1028 (1995).
- Rabinovitch, M. Pulmonary hypertension: pathophysiology as a basis for clinical decision making. *J. Heart Lung Transpl.* **18**, 1041–1053 (1999).
- Pinedo, H. M., Verheul, H. M., D'Amato, R. J. & Folkman, J. Involvement of platelets in tumour angiogenesis? *Lancet* **352**, 1775–1777 (1998).
- Seljelid, R., Jozefowski, S. & Sveinbjornsson, B. Tumor stroma. *Anticancer Res.* **19**, 4809–4822 (1999).
- Schaper, W. & Ito, W. D. Molecular mechanisms of coronary collateral vessel growth. *Circ. Res.* **79**, 911–919 (1996).
- Coussens, L. M. *et al.* Inflammatory mast cells up-regulate angiogenesis during squamous epithelial carcinogenesis. *Genes Dev.* **13**, 1382–1397 (1999).
- Heymans, S. *et al.* Inhibition of plasminogen activators or matrix metalloproteinases prevents cardiac rupture but impairs therapeutic angiogenesis and causes cardiac failure. *Nature Med.* **5**, 1135–1142 (1999).
- Carmeliet, P. & Collen, D. Development and disease in proteinase-deficient mice: role of the plasminogen, matrix metalloproteinase and coagulation system. *Thromb. Res.* **91**, 255–285 (1998).
- Buschmann, I. & Schaper, W. The pathophysiology of the collateral circulation (arteriogenesis). *J. Pathol.* **190**, 338–342 (2000).
- Silverman, K. J. *et al.* Angiogenic activity of adipose tissue. *Biochem. Biophys. Res. Commun.* **153**, 347–352 (1988).
- Sierra-Honigsmann, M. R. *et al.* Biological action of leptin as an angiogenic factor. *Science* **281**, 1683–1686 (1998).
- Isner, J. M. & Asahara, T. Angiogenesis and vasculogenesis as therapeutic strategies for postnatal neovascularization. *J. Clin. Invest.* **103**, 1231–1236 (1999).
- Carmeliet, P. *et al.* Impaired myocardial angiogenesis and ischemic cardiomyopathy in mice lacking the vascular endothelial growth factor isoforms VEGF164 and VEGF188. *Nature Med.* **5**, 495–502 (1999).
- Li, J. *et al.* PR39, a peptide regulator of angiogenesis. *Nature Med.* **6**, 49–55 (2000).
- Carmeliet, P. *et al.* Role of HIF-1 α in hypoxia-mediated apoptosis, cell proliferation and tumour

- angiogenesis. *Nature* **394**, 485–490 (1998).
68. Fidler, I. J. & Ellis, L. M. Chemotherapeutic drugs—more really is not better. *Nature Med.* **6**, 500–502 (2000).
69. Rohan, R. M., Fernandez, A., Udagawa, T., Yuan, J. & D'Amato, R. J. Genetic heterogeneity of angiogenesis in mice. *FASEB J.* **14**, 871–876 (2000).
70. Carmeliet, P. *et al.* Targeted deficiency or cytosolic truncation of the VE-cadherin gene in mice impairs VEGF-mediated endothelial survival and angiogenesis. *Cell* **98**, 147–157 (1999).
71. Bajou, K. *et al.* Absence of host plasminogen activator inhibitor 1 prevents cancer invasion and vascularization. *Nature Med.* **4**, 923–928 (1998).
72. Carmeliet, P. & Collen, D. Transgenic mouse models in angiogenesis and cardiovascular disease. *J. Pathol.* **190**, 387–405 (2000).
73. Jain, R. K., Schlenger, K., Höckel, M. & Yuan, F. Quantitative angiogenesis assays: progress and problems. *Nature Med.* **3**, 1203–1208 (1997).
74. Tsuzuki, Y. *et al.* VEGF modulation by targeting HIF-1 α /HRE/VEGF cascade differentially regulates vascular response and growth rate in tumors. *Cancer Res.* (in the press).
75. Leunig, M. *et al.* Angiogenesis, microvascular architecture, microhemodynamics, and interstitial fluid pressure during early growth of human adenocarcinoma LS174T in SCID mice. *Cancer Res.* **52**, 6553–6560 (1992).

Acknowledgements

We acknowledge the critical input of C. Mouta Carreira, B. Fenton, D. Fukumura, J. Samson, A. Kadambi, B. Stoll and E. diTomaso in manuscript preparation and L. L. Munn, M. Leunig and A. Vandenhoek in figure preparation. This work was supported by grants from the NIH, NSF, DOD, ACS, the National Foundation for Cancer Research and the Whitaker Foundation (R.K.J.), and from the VIB, FWO, Biomed and the VLK (P.C.). Owing to space limitations we have cited review articles instead of original references; relevant material that could not be included in this review is available at <http://steele.mgh.harvard.edu>.

Thrombin signalling and protease-activated receptors

Shaun R. Coughlin

Cardiovascular Research Institute and Departments of Medicine and Pharmacology, University of California at San Francisco, HSE-1300, 505 Parnassus Avenue, San Francisco, California 94143-0130, USA (e-mail: coughlin@cvrmail.ucsf.edu)

How does the coagulation protease thrombin regulate cellular behaviour? The protease-activated receptors (PARs) provide one answer. In concert with the coagulation cascade, these receptors provide an elegant mechanism linking mechanical information in the form of tissue injury or vascular leakage to cellular responses. Roles for PARs are beginning to emerge in haemostasis and thrombosis, inflammation, and perhaps even blood vessel development.

The serine protease thrombin regulates platelet aggregation, endothelial cell activation and other important responses in vascular biology. Thrombin's actions on cells raise an intriguing question. How does thrombin, a protease, act like a traditional hormone and elicit cellular responses? Understanding thrombin signalling will provide insight into haemostasis and inflammation, and, probably, embryonic development. Because thrombin and platelets have a central role in myocardial infarction and other pathological processes, understanding how thrombin activates platelets and other cells may suggest new strategies for therapy.

Protease-activated receptors (PARs) provide one answer to the question of how thrombin produces signals. PARs are G-protein-coupled receptors that use a fascinating mechanism to convert an extracellular proteolytic cleavage event into a transmembrane signal: these receptors carry their own ligands, which remain cryptic until unmasked by receptor cleavage. Recent advances in our understanding of PARs provide a working model for thrombin signalling in human platelets, reveal a surprising variation in the paradigm for PAR activation and evoke testable hypotheses regarding the roles of PARs in thrombosis and inflammation. It is therefore timely to review progress in our understanding of thrombin signalling and PARs in the context of vascular biology.

When and where is thrombin generated?

Thrombin is the main effector protease of the coagulation cascade, a series of zymogen conversions that is triggered when circulating coagulation factors contact tissue factor. Tissue factor is a type-I integral membrane protein that functions as an obligate cofactor for activation of zymogen factor X by factor VIIa. Factor Xa (with the assistance of cofactor factor Va) then converts prothrombin to active thrombin. Other zymogen conversions provide both amplification and negative feedback loops that regulate thrombin production. Thrombin is short lived in the circulation and, in the context of a normal endothelium, its actions tend to terminate its production. Thus thrombin is thought to act near the site at which it is produced^{1,2}.

Tissue factor is expressed by epithelial cells, macrophages and other cell types that are normally separated from blood and circulating coagulation factors. Classically, thrombin generation is triggered when

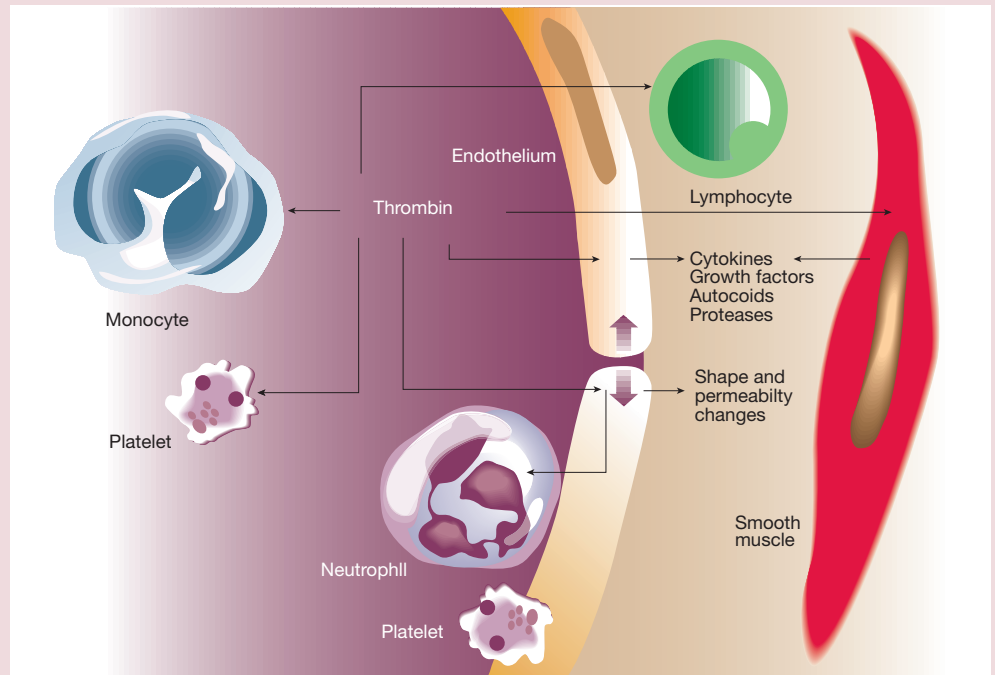
disruption of vascular integrity allows plasma coagulation factors to contact extravascular tissue factor. Thus the coagulation cascade provides a mechanism for converting mechanical information in the form of tissue damage and/or vascular leak into biochemical information in the form of the active protease thrombin.

Tissue factor is expressed at low levels on circulating monocytes and leukocyte-derived microparticles. These sources of intravascular tissue factor can be tethered to activated platelets and endothelial cells and concentrated in this way at sites of injury or inflammation^{3,4}. This alters the local balance between activation and inhibition of the coagulation cascade and triggers thrombin production. Tissue factor is also expressed at low levels by cytokine-stimulated endothelial cells, perhaps to promote thrombin generation at sites of inflammation⁵.

What are thrombin's actions on cells?

Thrombin converts circulating fibrinogen to fibrin monomer, which polymerizes to form fibrin, the fibrous matrix of blood clots. Thrombin also has a host of direct actions on cells⁶ (Fig. 1). It triggers shape change in platelets and the release of the platelet activators ADP, serotonin and thromboxane A₂, as well as chemokines and growth factors. It also mobilizes the adhesion molecule P-selectin and the CD40 ligand to the platelet surface^{7,8} and activates the integrin α IIb/ β 3 (ref. 9). The latter binds fibrinogen and von Willebrand factor (vWF) to mediate platelet aggregation¹. Thrombin also triggers expression of procoagulant activity on the platelet surface, which supports the generation of additional thrombin¹⁰. In cultured endothelial cells, thrombin causes release of vWF¹¹, the appearance of P-selectin at the plasma membrane¹¹, and production of chemokines — actions thought to trigger binding of platelets and leukocytes to the endothelial surface *in vivo*^{12,13}. Endothelial cells also change shape and endothelial monolayers show increased permeability in response to thrombin¹⁴ — actions predicted to promote local transudation of plasma proteins and oedema¹⁵. Thrombin can also regulate blood vessel diameter by endothelium-dependent vasodilation; in the absence of endothelium, thrombin acting on smooth muscle cells evokes vasoconstriction. In cultures of fibroblast or vascular smooth muscle cells, thrombin regulates cytokine production and is mitogenic, and in T lymphocytes it triggers calcium signalling and other responses. These cellular actions suggest that thrombin connects tissue

Figure 1 The actions of thrombin on blood cells and blood vessels. Thrombin is a multifunctional serine protease generated at sites of vascular injury. It is arguably the most effective agonist for platelet activation. Thrombin also elicits a host of responses in the vascular endothelium, including shape and permeability changes, mobilization of adhesive molecules to the endothelial surface and stimulation of autocrine (small molecule mediators such as prostaglandins and platelet-activating factor) and cytokine production. Thrombin is chemotactic for monocytes and mitogenic for lymphocytes and mesenchymal cells.



damage to both haemostatic and inflammatory responses and perhaps even to the decision to mount an immune response. They also raise the possibility that regulation of endothelial and other cell types by thrombin might have a role in leukocyte extravasation, vascular remodelling and/or angiogenesis in contexts other than tissue injury. The recent characterization of receptors that mediate thrombin signalling provides an opportunity to test these ideas.

How does thrombin talk to cells?

Thrombin signalling is mediated at least in part by a small family of G-protein-coupled PARs¹⁶. PAR1, the prototype of this family, is activated when thrombin cleaves its amino-terminal extracellular domain (exodomain) at a specific site^{17,18}. This cleavage unmasks a new N terminus that then serves as a tethered ligand, binding intramolecularly to the body of the receptor to effect transmembrane signalling¹⁷ (Fig. 2). Intermolecular ligation of PARs can occur but, not surprisingly, seems to be less efficient than intramolecular ligation^{19,20}. Synthetic peptides that mimic the tethered ligand of PAR1 activate the receptor independently of protease and receptor cleavage¹⁷. Thus PAR1 can be viewed as a peptide receptor that carries its own ligand. The latter remains silent until activated by cleavage of the PAR1 N-terminal exodomain. PAR1–thrombin interactions are accounted for by sequences surrounding the cleavage site within the N-terminal exodomain of the receptor, and cleavage at that site is both necessary and sufficient for PAR1 activation. Indeed, PAR1 mutants bearing enteropeptidase or trypsin cleavage sites in place of the thrombin cleavage site conferred the capacity for enteropeptidase or trypsin signalling, respectively, in heterologous expression systems. Thus the role of thrombin in PAR1 activation seems to be simply to unmask the receptor's tethered ligand^{6,16}.

PAR1 can couple to members of the $G_{12/13}$, G_q and G_i families and hence to a host of intracellular effectors (see Box 1). Such pluripotent signalling fits well with the known effects of thrombin on platelets, endothelial and other cells.

Irreversible activation and disposable receptors

The mechanism by which PAR1 is activated is striking in several ways. Cleavage of the receptor is irreversible, and the 'peptide agonist' unmasked by cleavage remains tethered to the receptor. Moreover, thrombin is an enzyme, implying that one thrombin molecule might

cleave and activate several molecules of PAR1. This raises several important and related questions. Given the irreversibility of the activation mechanism, how is PAR1 signalling terminated? Given that thrombin is an enzyme, how does PAR1 mediate responses that are dependent on thrombin concentration? And, given tethering of ligand to receptor, will development of drugs that block PAR1 signalling be possible? There are strong hints of interesting answers¹⁶.

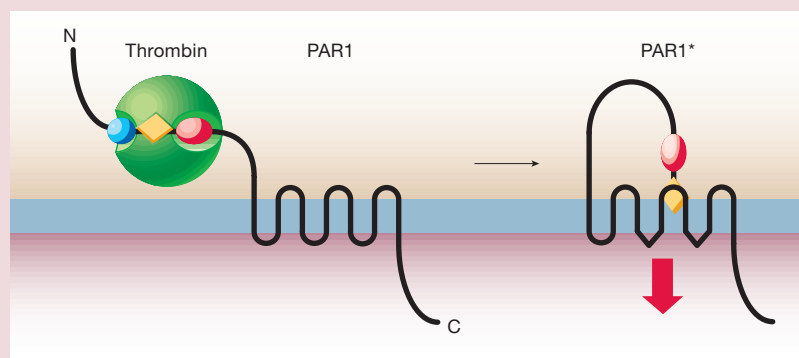
Like other G-protein-coupled receptors, activated PAR1 is rapidly uncoupled from signalling and internalized by phosphorylation-dependent mechanisms. Instead of recycling, it is then delivered to lysosomes for degradation with remarkable efficiency. Some PAR1 molecules that escape this fate appear to return to the cell surface with tethered ligand in an inactive state. Thus PAR1 is used once and then discarded. In fibroblasts and endothelial cells, responsiveness to thrombin is maintained by delivery of new PAR1 to the cell surface from a preformed intracellular pool. By contrast, in human megakaryocyte-like cell lines, recovery of PAR1 signalling requires new protein synthesis. Perhaps there is no need for a special resensitization mechanism in platelets. Once activated and incorporated into a clot, they are presumably not reused.

The rapid shut-off of activated PAR1 provides a plausible answer to how PAR1 mediates graded responses that vary with thrombin concentration²¹. Each cleaved receptor is active for a finite interval and therefore triggers production of some average 'unit' of second messenger (for example, inositol trisphosphate). Because the second messenger is itself cleared, the level of second messenger achieved is proportional to the rate at which receptors are cleaved and activated, and hence to thrombin concentration. Together with the relatively low avidity of the interaction between PAR1 and its tethered ligand, this mechanism makes us optimistic about the possibility of developing useful PAR1 antagonists. It suggests that, in order to attenuate cellular responses an antagonist need only delay PAR1 activation. Indeed, effective antagonists structurally related to the PAR1 tethered ligand have been generated²².

A family of PARs

Four PARs are known in mouse and human. Human PAR1 (refs 17, 18), PAR3 (ref. 23), and PAR4 (refs 24, 25) can be activated by thrombin. PAR2 is activated by trypsin²⁶ and tryptase²⁷ as well as by coagulation factors VIIa and Xa²⁸, but not by thrombin. It is certainly possible

Figure 2 Mechanism of PAR1 activation. Thrombin (large green sphere) recognizes the N-terminal exodomain of the G-protein-coupled thrombin receptor PAR1. This interaction uses sites both N-terminal (small blue sphere) and C-terminal (small pink oval) to the thrombin cleavage site. The latter sequence resembles the C-terminal tail of the thrombin inhibitor hirudin and binds to thrombin in an analogous manner. Thrombin cleaves the peptide bond between receptor residues Arg 41 and Ser 42. This serves to unmask a new N terminus, beginning with the sequence SFLLRN (diamond) that functions as a tethered ligand, docking intramolecularly with the body of the receptor to effect transmembrane signalling. Synthetic SFLLRN peptide, which mimics the tethered ligand sequence, will function as an agonist independently of receptor cleavage. Thus PAR1 is, in essence, a peptide receptor that carries its own ligand, the latter being active only after receptor cleavage.



that these receptors mediate responses to other proteases or even to peptide ligands *in vivo*. Indeed, cofactors that localize proteases to the cell surface and modulate their activity can help orchestrate PAR activation^{28,29}. Thus the full repertoire of proteases that signal through PARs remains to be defined.

It is worth noting that the N-terminal exodomains of PAR1 and PAR3 have thrombin-interacting sequences both N- and carboxy-terminal to the thrombin-cleavage site (Fig. 2). The C-terminal sequence resembles the C-terminal tail of the leech anticoagulant hirudin and, like the latter, binds to thrombin's fibrinogen-binding exosite; this interaction is important for receptor cleavage at low concentrations of thrombin⁶. The presence of such extended thrombin-interacting sequences in PAR1 and PAR3 is consistent with the notion that these receptors evolved to mediate responses to thrombin rather than to other proteases. A hirudin-like sequence is not evident in PAR4, and PAR4 indeed requires higher thrombin concentrations for activation than the other receptors^{24,25}.

PARs and platelet activation

Recent studies have provided a working model of thrombin signalling in human and mouse platelets and reveal both curious species differences and a variation on the paradigm for PAR activation. The model frames important questions regarding strategies for drug development and suggests that answers, at least in principle, can be derived from studies of PAR-knockout mice (Fig. 3).

Human platelets

Human platelets express PAR1 and PAR4, and activation of either is sufficient to trigger platelet secretion and aggregation^{17,24,30}. Antibodies to the thrombin-interaction site in PAR1 blocked receptor cleavage and platelet activation at low, but not high, concentrations of thrombin³⁰⁻³². By contrast, PAR4-blocking antibodies by themselves had no effect on platelet activation by thrombin, but when these were combined with PAR1 blockade, platelet activation was markedly inhibited, even at high concentrations of thrombin³⁰. These results suggest that PAR1 mediates activation of human platelets at low thrombin concentrations and that, in the absence of PAR1 function, PAR4 can mediate platelet activation but only at high thrombin concentrations (Fig. 3). Given that PAR1 does normally function in human platelets, what does PAR4 contribute? It is possible that PAR4 simply provides 'back up' in an important system. It is equally possible that PAR4, which lacks a thrombin-binding hirudin-like sequence, mediates responses to proteases other than thrombin. In this regard, platelet activation by cathepsin G³³, a granzyme released by activated neutrophils, seems to be mediated by PAR4 (ref. 34). PAR4 may make other unique contributions to platelet function. Indeed, PAR4 is activated and shut off more slowly than PAR1, and the tempo of calcium signalling in response to thrombin in human platelets appears to represent the sum contribution of both receptors³⁵.

It is worth noting that thrombin binds to the platelet surface

glycoprotein GPIb α ³⁶, part of a protein complex that also binds vWF and P-selectin³⁷. The role of this binding to GPIb α is unclear. It is possible that GPIb α serves as a cofactor that modulates thrombin's ability to cleave other platelet surface or plasma proteins, or that GPIb α has a more direct signalling role. Studies in knockout mice will soon reveal whether the known PARs account for thrombin signalling in platelets.

The presence of PAR1 and PAR4 in human platelets raises an important question regarding the development of antithrombotic drugs. Given that activation of PAR4 requires relatively high concentrations of thrombin, might inhibition of PAR1 be sufficient to prevent thrombosis? Or will inhibition of both PAR1 and PAR4 be required? In the absence of drugs that might be used to address this question in relevant animal models, answers in the near term are likely to come from PAR-deficient mice.

Mouse platelets

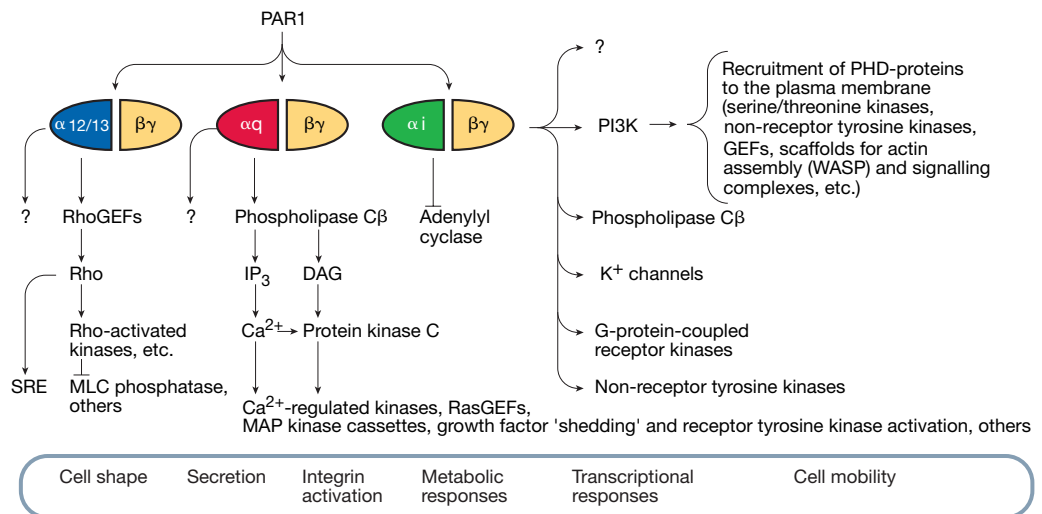
In contrast to human platelets, mouse platelets express PAR3 and PAR4 (ref. 25). Indeed, PAR1-activating peptides activate human but not murine platelets³⁸⁻⁴⁰, and knockout of mouse PAR1 (mPAR1) had no effect on thrombin signalling in mouse platelets but abolished thrombin signalling in fibroblasts⁴⁰. These observations triggered a search for other thrombin receptors in mouse platelets and led to the identification of PAR3 (ref. 23). Expression of human PAR3 cDNA in COS cells or *Xenopus* oocytes conferred phosphoinositide hydrolysis in response to low concentrations of thrombin, and *in situ* hybridization using a mouse PAR3 probe detected mPAR3 mRNA in mouse megakaryocytes²³. Knockout of mouse PAR3 revealed PAR3 to be necessary for activation of mouse platelets at low but not high concentrations of thrombin. Persistent thrombin signalling in PAR3-deficient mouse platelets was attributable to mPAR4 (ref. 25). On the face of it, these data conjured up a dual-receptor model analogous to that described for human platelets. In mouse platelets, PAR3 mediated activation at low thrombin concentrations and, in the absence of PAR3 function, PAR4 triggered activation at high thrombin concentrations²⁵.

Subsequent characterization of the mouse homologue of PAR3 presented a paradox. In spite of strong evidence that mPAR3 was necessary for mouse platelet responses to low concentrations of thrombin, expression of mPAR3 cDNA in heterologous expression systems failed to confer the property of thrombin signalling. Resolution of this paradox came in the form of an interesting variation on the mechanism of PAR activation²⁹. Whereas expression of mPAR3 in COS cells did not, by itself, confer thrombin signalling, co-expression of mPAR3 with mPAR4 reliably enhanced both mPAR4 cleavage and signalling at low concentrations of thrombin compared with mPAR4 alone. When tethered to the plasma membrane, the N-terminal exodomain of mPAR3 was sufficient for this activity, and the thrombin-interacting sequences within this domain were necessary. Thus, it appears that mPAR3 does not by itself mediate transmembrane signalling, but instead functions as a cofactor for cleavage and

Box 1

Thrombin receptor signalling

PAR1 can couple to members of the $G_{12/13}$, G_q , and G_i families^{45–47} to impact on a substantial network of signalling pathways, as shown in the figure. The α -subunits of G_{12} and G_{13} bind RhoGEFs (guanine-nucleotide exchange factors, which activate small G proteins such as Rho)^{48–50}, providing a pathway to Rho-dependent cytoskeletal responses that are likely to be involved in shape changes in platelets⁵¹ and permeability and migration in endothelial cells^{52,53}. G_{α_q} activates phospholipase $C\beta$ ⁵⁴, triggering phosphoinositide hydrolysis which results in calcium mobilization and activation of protein kinase C. This provides a pathway to calcium-regulated kinases and phosphatases, GEFs, mitogen-activated protein (MAP) kinase cassettes, and other proteins that mediate cellular responses ranging from granule secretion, integrin activation and aggregation in platelets⁵⁵, to transcriptional responses in endothelial and mesenchymal cells. G_{α_i} inhibits adenylate cyclase, an action known to promote platelet responses. $G\beta\gamma$ subunits can activate phosphoinositide 3-kinase (PI(3)K)⁵⁶ and other lipid-modifying enzymes, protein kinases and ion channels⁵⁷. PI(3)K modifies the inner leaflet of the plasma membrane to provide attachment sites for a host of signalling proteins⁵⁸. PAR1 activation can also activate cell-surface 'sheddases' which liberate ligands for receptor tyrosine kinases, providing a link between thrombin and receptors involved in cell growth and differentiation⁵⁹. The pleiotropic effects of PAR1 activation are consistent with many of thrombin's diverse actions on cells. IP₃, inositol trisphosphate; DAG, diacylglycerol; SRE, serum response element; PHD, pleckstrin homology domain.



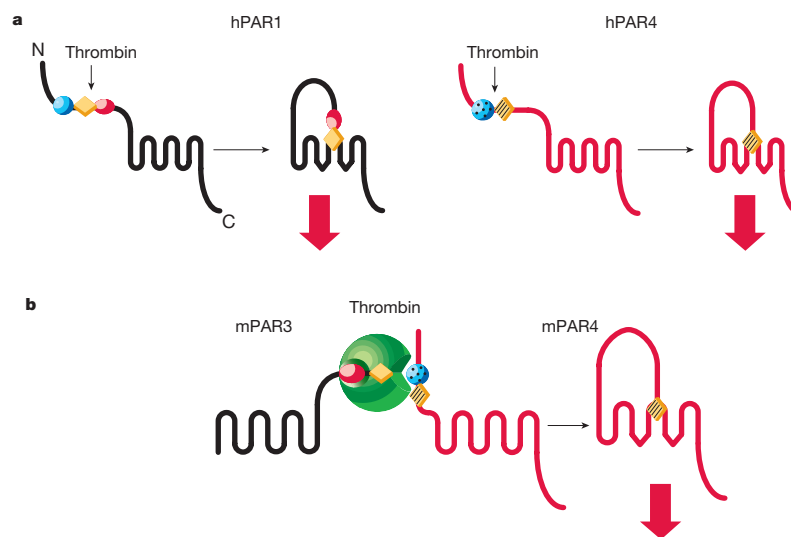
activation of mPAR4 at low thrombin concentrations — a curious form of G-protein-coupled receptor interaction in which one receptor acts as an accessory protein that aids 'ligation' of another (Fig. 3)²⁹. This model predicts that thrombin signalling in mouse platelets is dependent on PAR4. A definitive test of this prediction will be possible with platelets from PAR4-deficient mice, which should be unresponsive to thrombin despite the presence of mPAR3 (Fig. 3). Whether mPAR3 and mPAR4 heterodimerize, and whether other similar PAR–PAR interactions will be found, is not known. There is

no evidence to suggest an analogous interaction between human PAR1 (hPAR1) and hPAR4 or between the thrombin-binding site GPIbα in human platelets and hPAR4.

Utility of mouse models

Despite species differences, mouse models may provide important hints on how to inhibit thrombin signalling in human platelets. The model in Fig. 3 makes several predictions. First, PAR3-deficient mouse platelets are analogous to PAR1-inhibited human platelets — both rely on PAR4 for thrombin signalling. If PAR3 deficiency protects

Figure 3 Thrombin signalling in human and mouse platelets. Human platelets express PAR1 and PAR4, and available data suggest that these receptors can independently mediate thrombin signalling — PAR1 at low and PAR4 at high thrombin concentrations. By contrast, mouse platelets express PAR3 and PAR4 and, surprisingly, it seems that mPAR3, rather than itself mediating transmembrane signalling, functions as a cofactor that supports cleavage and activation of mPAR4 at low thrombin concentrations.



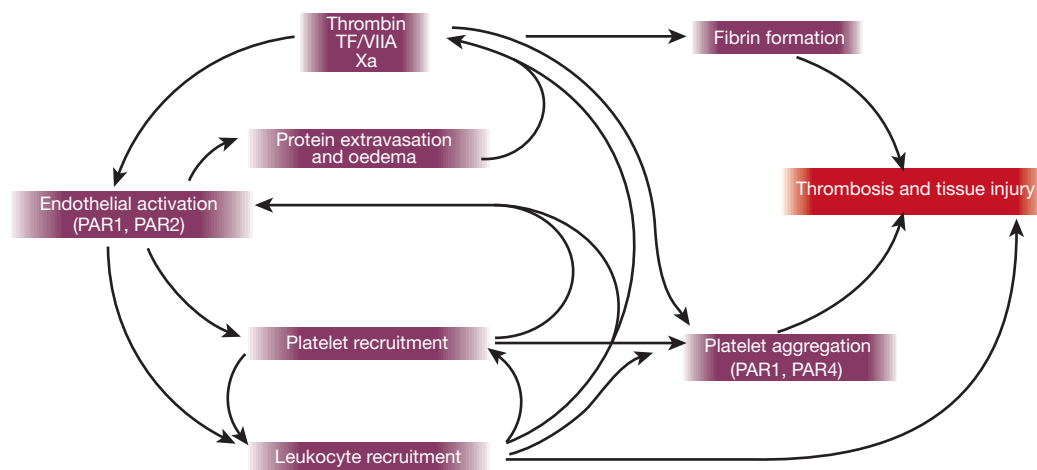
Box 2

Potential roles for PARs in disease

Thrombosis of the arteries that supply the heart, brain and other vital organs is a major cause of morbidity and mortality. Both thrombin and platelets are clearly important in acute arterial thrombosis and, given the remarkable effectiveness of thrombin as a platelet agonist, it is reasonable to postulate an important role for platelet activation by thrombin. Thrombin has many actions, however, and platelets respond to multiple agonists. The relative importance of thrombin signalling in platelets in the complex interplay among platelet, plasma and vessel wall factors in thrombosis is still to be determined.

Less widely appreciated is the potential role of signalling by coagulation proteases in inflammatory processes. As described by Esmon⁶⁰, molecular links between coagulation and inflammation have been established, and coagulation inhibitors are effective in primate models of septic shock. Indeed, Eli Lilly recently announced that activated protein C, an important negative regulator of thrombin generation⁶¹, is efficacious in septic shock in humans. Might PARs participate in the link between the coagulation cascade and inflammation? A positive feedback loop like that shown in the figure may contribute to the extraordinary leukocyte activation, disseminated intravascular coagulation, and microvascular thrombosis and haemorrhagic infarction (purpura fulminans) seen in sepsis^{2,60}.

Thrombin activates endothelial PAR1, and factor Xa, and perhaps tissue factor/factor VIIa complex (TF/VIIa), activate endothelial PAR2. PAR signalling upregulates adhesion molecules on the endothelial surface and triggers production of autocoagulants and chemokines that activate neutrophils and monocytes (see main text). This leads to binding, rolling, and eventual attachment of platelets and leukocytes to the endothelial surface. These local concentrations of



leukocytes and microparticles bearing tissue factor^{3,4}, along with platelet procoagulant activity¹⁰, may trigger further thrombin generation⁶². Thrombin also increases the permeability of the endothelium, and PAR activation triggers oedema formation, at least in part by triggering mast-cell degranulation¹⁵. This may promote generation of additional thrombin as plasma coagulation factors contact extravascular tissue factor. Platelets and leukocytes can directly activate endothelial cells by presenting CD40 ligand and other mediators, upregulating not only adhesion molecules and cytokines⁸ but also tissue factor and PAR2 (refs 5, 63). Leukocytes and platelets can themselves interact via P-selectin⁶², and neutrophils can activate platelets by release of cathepsin G. Ultimately, leukocyte products may directly injure tissues, and thrombin may trigger fibrin formation, platelet aggregation, microvascular thrombosis and, potentially, tissue ischaemia and infarction. Such undamped positive feedback between coagulation and inflammation may be made more likely by genetic deficiencies in natural anticoagulant pathways⁶⁴. Clearly, a host of cell types and signalling systems orchestrate inflammatory responses, and the relative importance of PARs in sepsis and in less dramatic inflammatory processes is unknown. The recent observation that PAR1 deficiency is protective in a mouse model of antibody-mediated glomerulonephritis⁴³ supports the notion that signalling by coagulation proteases may contribute to inflammatory responses.

against thrombosis in mouse models, PAR1 inhibition may be worth investigating as an antithrombotic strategy in humans. Second, PAR4-deficient mouse platelets may be analogous to human platelets in which both PAR1 and PAR4 function are blocked — thrombin signalling should be absent in both. Thus a PAR4-deficient mouse might provide an opportunity to define the importance of thrombin-triggered platelet activation in haemostasis and thrombosis.

PARs in endothelial activation

PAR1 seems to be the major mediator of thrombin signalling in vascular endothelial cells in both mice and humans, and most of the actions of thrombin on endothelial cells described above have been reproduced using PAR1 agonist peptide. Endothelial cells also express PAR2, which may mediate responses to tryptase released from mast cells²⁷ or to coagulation factors VIIa or Xa²⁸ in this setting.

What are the functions of endothelial PARs? One might imagine the following scenario. Tissue injury, whether by trauma, infection or metabolic or inflammatory mediators, triggers local generation of coagulation proteases and/or release of mast-cell tryptase which, by way of PARs, activate endothelial cells (Fig. 1). The activated endothelial surface in turn promotes adhesion and rolling of

platelets and leukocytes as well as leakage of plasma proteins to the extravascular space. Thrombin also triggers endothelial production of platelet-activating factor, a potent neutrophil activator⁴¹, as well as the interleukins IL-6 and IL-8 (ref. 42). Thus PARs may link tissue injury to endothelial responses that recruit platelets, leukocytes and effector proteins to examine the locale for damage or infection.

The possibility of a positive feedback loop in which thrombin triggers endothelial responses that beget additional thrombin generation and endothelial activation is clear (Box 2). Undamped, such a system would trigger intravascular thrombosis and, perhaps, local tissue damage from leukocyte products. On a microscopic scale, this might be beneficial for walling off infection. However, disseminated intravascular coagulation with microvascular thrombosis and tissue infarction can occur in the setting of a strong systemic inflammatory stimulus (for example, sepsis) and/or deficiencies that disinhibit thrombin production (for example, protein C deficiency, protein S deficiency and the presence of factor V Leiden).

PAR-deficient mice provide an opportunity to test the role of endothelial PARs in inflammatory responses. The species differences in PAR expression between mouse and human may be fortunate in this regard. PAR1 appears to be the major thrombin receptor in

endothelial cells in both species. Because PAR1 is expressed in human platelets but not in those of mice, however, PAR1-deficient mice offer an opportunity to abolish thrombin signalling in endothelial cells without perturbing platelet signalling. This can help define the contribution of endothelial activation by thrombin to thrombosis and inflammation. Intriguingly, PAR1 deficiency did protect against leukocyte infiltration and renal damage in a mouse model of antibody-mediated glomerulonephritis⁴³.

Thrombin signalling in embryonic development

Approximately 50% of PAR1-deficient mouse embryos die at mid-gestation⁴⁰. PAR1 does not act in mouse platelets, so this and other recent studies in knockout mice suggest that signalling by coagulation proteases and PARs may have an important role in embryonic development that is unrelated to haemostasis in any usual sense^{16,44}. The available data suggest that PAR1 and coagulation factors may contribute to normal blood vessel development. This is exciting, in that it may point to a new role for the 'coagulation' cascade — one of monitoring and regulating new blood vessel formation.

Future directions

The studies described above raise a host of questions regarding the molecular mechanisms of PAR activation and protease signalling. How general are PAR–PAR interactions and is receptor oligomerization involved? To what extent do cofactors increase the diversity of proteases to which cells can respond through PARs? Will the known PARs account completely for signalling by thrombin and other coagulation proteases, or will new PARs and/or other mechanisms be identified? Because PAR1 and PAR4 are the only PARs known to mediate transmembrane signalling in response to thrombin in the mouse, the presence or absence of residual thrombin signalling in cells from mice deficient in both PAR1 and PAR4 will be telling.

Important questions also remain regarding the roles of PARs in physiology and disease. For example, thrombin is a powerful activator of platelets and it is clear that both thrombin and platelets are important for haemostasis and thrombosis. But in addition to activating platelets, thrombin triggers fibrin formation, and platelets can be activated by a host of other mechanisms. Thus the relative importance of thrombin activation of platelets in haemostasis and thrombosis is unknown. As discussed above, the phenotype of a PAR4-knockout mouse may be enlightening in this respect. Similarly, a panoply of signalling systems and cell types orchestrates inflammatory responses, and efforts to define the relative contribution of PARs are just beginning.

The answers to these questions will influence decisions as to whether or not PARs are rational drug targets. Blockade of platelet activation by thrombin might well be a useful antithrombotic strategy. Attenuating inflammatory responses by blocking PAR signalling in endothelial cells is a more novel and untested notion, and affecting new blood vessel formation by the same route is more speculative still. Results from mouse models may stimulate the development of drugs to further explore these ideas. □

- Colman, R. W., Marder, V. J., Salzman, E. W. & Hirsh, J. in *Hemostasis and Thrombosis* (eds Colman, R. W., Marder, V. J., Salzman, E. W. & Hirsh, J.) 3–18 (Lippincott, Philadelphia, 1994).
- Esmon, C. T. *et al.* Inflammation, sepsis, and coagulation. *Haematologica* **84**, 254–259 (1999).
- Osterud, B. Tissue factor expression by monocytes: regulation and pathophysiological roles. *Blood Coag. Fibrin* **9**(Suppl. 1), S9–S14 (1998).
- Giesen, P. L. *et al.* Blood-borne tissue factor: another view of thrombosis. *Proc. Natl Acad. Sci. USA* **96**, 2311–2315 (1999).
- Bevilacqua, M. P. & Gimbrone, M. A. Jr Inducible endothelial functions in inflammation and coagulation. *Semin. Thromb. Hemost.* **13**, 425–433 (1987).
- Coughlin, S. R. Sol Sherry lecture in thrombosis: how thrombin 'talks' to cells: molecular mechanisms and roles *in vivo*. *Arterioscl. Thromb. Vasc. Biol.* **18**, 514–518 (1998).
- Stenberg, P. E., McEver, R. P., Shuman, M. A., Jacques, Y. V. & Bainton, D. F. A platelet alpha-granule membrane protein (GMP-140) is expressed on the plasma membrane after activation. *J. Cell Biol.* **101**, 880–886 (1985).
- Henn, V. *et al.* CD40 ligand on activated platelets triggers an inflammatory reaction of endothelial cells. *Nature* **391**, 591–594 (1998).
- Hughes, P. E. & Pfaff, M. Integrin affinity modulation. *Trends Cell Biol.* **8**, 359–364 (1998).
- Sims, P. J., Wiedmer, T., Esmon, C. T., Weiss, H. J. & Shattil, S. J. Assembly of the platelet

- prothrombinase complex is linked to vesiculation of the platelet plasma membrane. *J. Biol. Chem.* **264**, 17049–17057 (1989).
- Hattori, R., Hamilton, K. K., Fugate, R. D., McEver, R. P. & Sims, P. J. Stimulated secretion of endothelial vWF is accompanied by rapid redistribution to the cell surface of the intracellular granule membrane protein GMP-140. *J. Biol. Chem.* **264**, 7768–7771 (1989).
- Subramaniam, M. *et al.* Defects in hemostasis in P-selectin-deficient mice. *Blood* **87**, 1238–1242 (1996).
- Frenette, P. S., Mayadas, T. N., Rayburn, H., Hynes, R. O. & Wagner, D. D. Susceptibility to infection and altered hematopoiesis in mice deficient in both P- and E-selectins. *Cell* **84**, 563–574 (1996).
- Lum, H. & Malik, A. B. Regulation of vascular endothelial barrier function. *Am. J. Physiol.* **267**, L223–L241 (1994).
- Cirino, G. *et al.* Thrombin functions as an inflammatory mediator through activation of its receptor. *J. Exp. Med.* **183**, 821–827 (1996).
- Coughlin, S. R. How the protease thrombin talks to cells. *Proc. Natl Acad. Sci. USA* **96**, 11023–11027 (1999).
- Vu, T.-K. H., Hung, D. T., Wheaton, V. I. & Coughlin, S. R. Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* **64**, 1057–1068 (1991).
- Rasmussen, U. B. *et al.* cDNA cloning and expression of a hamster alpha-thrombin receptor coupled to Ca²⁺ mobilization. *FEBS Lett.* **288**, 123–128 (1991).
- Chen, J., Ishii, M., Wang, L., Ishii, K. & Coughlin, S. R. Thrombin receptor activation: confirmation of the intramolecular tethered ligand hypothesis and discovery of an alternative intermolecular ligand mode. *J. Biol. Chem.* **269**, 16041–16045 (1994).
- O'Brien, P. J. *et al.* Thrombin responses in human endothelial cells. Contributions from receptors other than PAR1 include the transactivation of PAR2 by thrombin-cleaved PAR1. *J. Biol. Chem.* **275**, 13502–13509 (2000).
- Ishii, K., Hein, L., Kobilka, B. & Coughlin, S. R. Kinetics of thrombin receptor cleavage on intact cells. Relation to signaling. *J. Biol. Chem.* **268**, 9780–9786 (1993).
- Bernatowicz, M. S. *et al.* Development of potent thrombin receptor antagonist peptides. *J. Med. Chem.* **39**, 4879–4887 (1996).
- Ishihara, H. *et al.* Protease-activated receptor 3 is a second thrombin receptor in humans. *Nature* **386**, 502–506 (1997).
- Xu, W. F. *et al.* Cloning and characterization of human protease-activated receptor 4. *Proc. Natl Acad. Sci. USA* **95**, 6642–6646 (1998).
- Kahn, M. L. *et al.* A dual thrombin receptor system for platelet activation. *Nature* **394**, 690–694 (1998).
- Nystedt, S., Emilsson, K., Wahlestedt, C. & Sundelin, J. Molecular cloning of a potential novel proteinase activated receptor. *Proc. Natl Acad. Sci. USA* **91**, 9208–9212 (1994).
- Molino, M. *et al.* Interactions of mast cell tryptase with thrombin receptors and PAR-2. *J. Biol. Chem.* **272**, 4043–4049 (1997).
- Camerer, E., Huang, W. & Coughlin, S. R. Tissue factor- and Factor X-dependent activation of PAR2 by Factor VIIa. *Proc. Natl Acad. Sci. USA* **97**, 5255–5260 (2000).
- Nakanishi-Matsui, M. *et al.* PAR3 is a cofactor for PAR4 activation by thrombin. *Nature* **404**, 609–613 (2000).
- Kahn, M. L., Nakanishi-Matsui, M., Shapiro, M. J., Ishihara, H. & Coughlin, S. R. Protease-activated receptors 1 and 4 mediate activation of human platelets by thrombin. *J. Clin. Invest.* **103**, 879–887 (1999).
- Hung, D. T., Vu, T. K., Wheaton, V. I., Ishii, K. & Coughlin, S. R. Cloned platelet thrombin receptor is necessary for thrombin-induced platelet activation. *J. Clin. Invest.* **89**, 1350–1353 (1992).
- Brass, L. F. *et al.* Structure and function of the human platelet thrombin receptor. Studies using monoclonal antibodies directed against a defined domain within the receptor N terminus. *J. Biol. Chem.* **267**, 13795–13798 (1992).
- Selak, M. A., Chignard, M. & Smith, J. B. Cathepsin G is a strong platelet agonist released by neutrophils. *Biochem. J.* **251**, 293–299 (1988).
- Sambrano, G. R. *et al.* Cathepsin G activates protease-activated receptor-4 in human platelets. *J. Biol. Chem.* **275**, 6819–6823 (2000).
- Shapiro, M. J., Weiss, E. J., Faruqi, T. R. & Coughlin, S. R. Protease-activated receptors 1 and 4 are shut off with distinct tempos after activation by thrombin. *J. Biol. Chem.* (in the press).
- Okamura, T., Hasitz, M. & Jamieson, G. A. Platelet glycoprotein: interaction with thrombin and role as a thrombin receptor on the platelet surface. *J. Biol. Chem.* **253**, 3435–3443 (1978).
- Andrews, R. K. *et al.* The glycoprotein Ib-IX-V complex in platelet adhesion and signaling. *Thromb. Haemost.* **82**, 357–364 (1999).
- Connolly, T. M. *et al.* Species variability in platelet and other cellular responsiveness to thrombin receptor-derived peptides. *Thromb. Haemost.* **72**, 627–633 (1994).
- Derian, C. K., Santulli, R. J., Tomko, K. A., Haertlein, B. J. & Andrade-Gordon, P. Species differences in platelet responses to thrombin and SFLRN. Receptor-mediated calcium mobilization and aggregation and regulation by protein kinases. *Thromb. Res.* **6**, 505–519 (1995).
- Connolly, A. J., Ishihara, H., Kahn, M. L., Farese, R. V. & Coughlin, S. R. Role of the thrombin receptor in development and evidence for a second receptor. *Nature* **381**, 516–519 (1996).
- Zimmerman, G. A. *et al.* Platelet-activating factor (PAF): signalling and adhesion in cell-cell interactions. *Adv. Exp. Med. Biol.* **416**, 297–304 (1996).
- Johnson, K. *et al.* Potential mechanisms for a proinflammatory vascular cytokine response to coagulation activation. *J. Immunol.* **160**, 5130–5135 (1998).
- Cunningham, M. A. *et al.* Protease-activated receptor 1 mediates thrombin-dependent, cell-mediated renal inflammation in crescentic glomerulonephritis. *J. Exp. Med.* **191**, 455–462 (2000).
- Bugge, T. H., *et al.* Fatal embryonic bleeding events in mice lacking tissue factor, the cell-associated initiator of blood coagulation. *Proc. Natl Acad. Sci. USA* **93**, 6258–6263 (1996).
- Hung, D. T., Wong, Y. H., Vu, T.-K. H. & Coughlin, S. R. The cloned platelet thrombin receptor couples to at least two distinct effectors to stimulate both phosphoinositide hydrolysis and inhibit adenylyl cyclase. *J. Biol. Chem.* **353**, 20831–20834 (1992).
- Offermanns, S., Laugwitz, K.-L., Spicher, K. & Schultz, G. G proteins of the G12 family are activated via thromboxane A2 and thrombin receptors in human platelets. *Proc. Natl Acad. Sci. USA* **91**, 504–508 (1994).
- Barr, A. J., Brass, L. F. & Manning, D. R. Reconstitution of receptors and GTP-binding regulatory proteins (G proteins) in Sf9 cells. A direct evaluation of selectivity in receptor–G protein coupling. *J.*

- Biol. Chem.* **272**, 2223–2229 (1997).
48. Kozasa, T. *et al.* p115 RhoGEF, a GTPase-activating protein for G α 12 and G α 13. *Science* **280**, 2109–2111 (1998).
 49. Hart, M. J. *et al.* Direct stimulation of the guanine nucleotide exchange activity of p115 RhoGEF by G α 13. *Science* **280**, 2112–2114 (1998).
 50. Fukuhara, S., Murga, C., Zohar, M., Igishi, T. & Gutkind, J. S. A novel PDZ domain containing guanine nucleotide exchange factor links heterotrimeric G proteins to Rho. *J. Biol. Chem.* **274**, 5868–5879 (1999).
 51. Klages, B., Brandt, U., Simon, M. I., Schultz, G. & Offermanns, S. Activation of G12/G13 results in shape change and rho/rho kinase-mediated myosin light chain phosphorylation in mouse platelets. *J. Cell. Biol.* **144**, 745–754 (1999).
 52. Vouret-Craviari, V., Boquet, P., Pouyssegur, J. & Van Obberghen-Schilling, E. Regulation of the actin cytoskeleton by thrombin in human endothelial cells: role of Rho proteins in endothelial barrier function. *Mol. Biol. Cell* **9**, 2639–2653 (1998).
 53. Offermanns, S., Mancino, V., Revel, J.-P. & Simon, M. I. Vascular system defects and impaired cell chemokinesis as a result of G α 13 deficiency. *Science* **275**, 533–536 (1997).
 54. Taylor, S., Chae, H. Z., Rhee, S.-G. & Exton, J. H. Activation of the B1 isozyme of phospholipase C by α subunits of the Gq class of G proteins. *Nature* **350**, 516–518 (1991).
 55. Offermanns, S., Toombs, C. F., Hu, Y. H. & Simon, M. I. Defective platelet activation in G α_q -deficient mice. *Nature* **389**, 183–186 (1997).
 56. Stoyanov, B. *et al.* Cloning and characterization of a G protein-activated human phosphoinositide-3 kinase. *Science* **269**, 690–693 (1995).
 57. Clapham, D. E. & Neer, E. J. G protein beta gamma subunits. *Annu. Rev. Pharmacol. Toxicol.* **37**, 167–203 (1997).
 58. Leervers, S. J., Vanhaesebroeck, B. & Waterfield, M. D. Signalling through phosphoinositide 3-kinases: the lipids take centre stage. *Curr. Opin. Cell Biol.* **11**, 219–225 (1999).
 59. Prenzel, N. *et al.* EGF receptor transactivation by G-protein-coupled receptors requires metalloproteinase cleavage of proHB-EGF. *Nature* **402**, 884–888 (1999).
 60. Esmon, C. T. Introduction: are natural anticoagulants candidates for modulating the inflammatory response to endotoxin? *Blood* **95**, 1113–1116 (2000).
 61. Esmon, C. T. *et al.* Regulation and functions of the protein C anticoagulant pathway. *Haematologica* **84**, 363–368 (1999).
 62. Palabrica, T. *et al.* Leukocyte accumulation promoting fibrin deposition is mediated *in vivo* by P-selectin on adherent platelets. *Nature* **359**, 848–851 (1992).
 63. Nystedt, S., Ramakrishnan, V. & Sundelin, J. The proteinase-activated receptor 2 is induced by inflammatory mediators in human endothelial cells. Comparison with the thrombin receptor. *J. Biol. Chem.* **271**, 14910–14915 (1996).
 64. Inbal, A. *et al.* Purpura fulminans induced by disseminated intravascular coagulation following infection in 2 unrelated children with double heterozygosity for factor V Leiden and protein S deficiency. *Thromb. Haemost.* **77**, 1086–1089 (1997).

Acknowledgements

I thank B. Black, H. Bourne, I. Charo, P.-T. Chuang, C. Esmon and the members of my laboratory for critical reading of the manuscript, and T. Schoop for illustrations.



Perspectives for vascular genomics

Edward M. Rubin* & Alan Tall†

*Lawrence Berkeley National Laboratory, Genome Sciences Department, 1 Cyclotron Road, Berkeley, California 94720, USA (e-mail: emrubin@lbl.gov)

†Division of Molecular Medicine, Department of Medicine, Columbia University, College of Physicians and Surgeons, New York, New York 10032, USA

Diseases of the vascular system result from a complex mixture of genetic and environmental factors. Data sets, technologies and strategies emanating from the human genome programme have been applied to the analysis of both rare single-gene and common multigenic vascular disorders. Genomic approaches including inter- and intraspecies sequence comparisons, genotyping with dense marker sets spanning the genome, large-scale mutagenesis screens of model organisms, and genome-wide expression profiling have all begun to contribute to the identification of new genes and mechanisms that are central to cardiovascular disease processes.

Although many of the genes responsible for rare single-gene vascular disorders have been described, the vast majority of genes that contribute to common polygenic vascular conditions remain unidentified. Just as the application of recombinant molecule technologies to single-gene disorders markedly accelerated their elucidation, it is likely that the application of recently developed genomic technologies will contribute to the deciphering of genetically complex vascular disorders. Several genomic strategies that have matured through the extensive analysis of simple model organisms are now being applied to the study of this complex mixture of human diseases. Here we discuss ways in which genomic technologies and resources are providing new insights into vascular biology.

Genomic sequencing and gene regulation

The ability to rapidly generate large amounts of genomic sequence is one of the most recent and spectacular advances of the human genome programme. The number of DNA bases sequenced in the year 2000 already far exceeds the combined yearly total of the previous 20 years since the dawn of DNA sequencing^{1,2}. The increased capabilities for generating genomic sequence are already being exploited to decipher issues that are relevant to vascular biology with regard to gene regulation through large-scale sequence

comparisons both between and within species.

Although many common disorders of the cardiovascular system are believed to be due to alterations in gene expression, relatively few sequences regulating gene expression have been identified. This is in part due to the classical, labour-intensive molecular approaches required to identify such sequences. In contrast, interspecies comparative sequence analyses offer a high-throughput means to identify sequences that confer specific function, including those with gene regulatory activity³. Selective pressure to maintain these sequences, resisting the general divergent drift of evolution, occurs in non-coding sequences that are recognized by gene regulatory factors in a similar way to that occurring in coding sequences specifying important structural domains of proteins. In response to the growing availability of genomic sequence, interspecies sequence comparison is increasingly being used to identify those non-coding sequences with gene regulatory activity.

Large-scale mouse/human sequence-based analyses targeted at identifying the regulatory elements for genes involved in vascular biology have included the analysis of a megabase interval containing a clustering of interleukin (IL) genes³ and the 320-kilobase (kb) region surrounding the stem-cell leukaemia (SCL) gene⁴. Using an interspecies sequence comparison strategy to identify non-coding regulatory sequences, Loots *et al.*³ investigated an orthologous interval on mouse chromosome 11 and the long arm (q) of

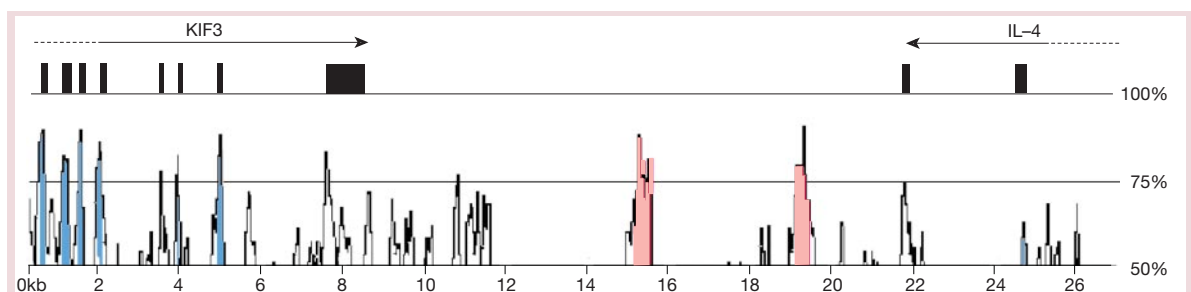


Figure 1 Comparative sequences of human and mouse DNA. In this VISTA display⁵, the horizontal axis represents 26 kb of human sequence while the vertical axis indicates the percentage identity between this and the orthologous mouse sequence. Human/mouse sequences that are >70% identical over >100 bp are indicated by pink vertical bars when they are present within non-coding sequences and blue vertical bars when within coding sequences. The comparative analysis easily identifies most exons in the highly conserved *KIF3* gene, although this is not the case with *IL-4*, a gene minimally conserved between mice and humans. Two conserved, non-coding sequences indicated by the vertical red bars 3' to *IL-4* gene show a significantly higher degree of human/mouse sequence conservation than the exons of *IL-4*. (See <http://www-gsd.lbl.gov/vista> for more information on VISTA.)

human chromosome 5 at band 31 (5q31) containing a total of 23 genes of which 5 were cytokines that participate in the inflammatory response. Of the many highly conserved non-coding sequences identified, analysis of the largest element (82% identity over 400 base pairs (bp)) in transgenic mice showed it to regulate the expression of several genes spread over 120 kb, both upstream (*IL-4*) and downstream (*IL-5* and *IL-13*) of the element. (The VISTA (visualization tool for alignment) plot⁵ displayed in Fig. 1 illustrates the ease with which conserved non-coding sequences can be identified through mouse/human comparisons; see also <http://www.gsd.lbl.gov/vista>.)

A similar strategy was used to identify sequences regulating the expression of *SCL*, a gene that is crucial in vasculogenesis and haematopoiesis^{6,7}. Human/mouse comparative sequence analysis of this region identified several new, conserved non-coding sequences, one of which was shown to regulate *SCL* expression in the brain. The interleukin gene cluster and the *SCL* studies highlight the complexity of long-range regulatory elements and the power of comparative biology in discovering and deciphering the properties of such elements.

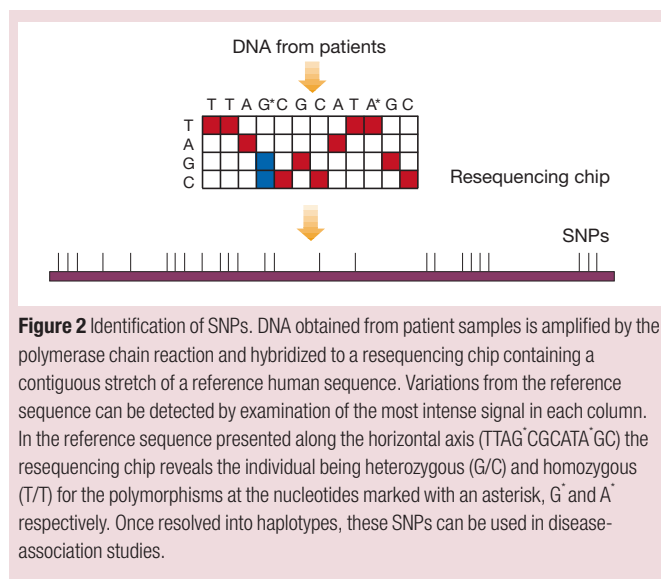
Sequence polymorphisms and complex disorders

Sequence polymorphism and gene mapping

A variety of polymorphic DNA markers localized throughout the human genome have been used extensively to map genes contributing to vascular disorders. This has been most successful in the positional cloning of genes contributing to rare mendelian disorders affecting the vasculature, ranging from hypertension⁸ to atherosclerosis^{9–11}. By following the familial segregation of short tandem-repeat polymorphisms with disease phenotypes, the fine mapping of disease genes and subsequent identification of causative mutations has become a straightforward process. In contrast, identifying genes that contribute to common complex disorders of the vasculature remains a significant challenge because, unlike most mendelian disorders, they result from the interactions of several genes. An example of this is hypertension, where several genes acting in concert determine an individual's blood pressure. Blood pressure levels are a quantitative trait continuously distributed throughout a population. Patients whose levels are at the upper extreme of the blood pressure distribution are defined as being hypertensive. The genetic interval, usually large, to which a gene contributing to a quantitative trait has been localized is called a quantitative trait locus (QTL). Through association studies QTLs for many complex disorders of vasculature have been mapped. The identification of the responsible gene(s) within the QTL genetic interval is the principal hurdle presently confronting the study of common complex disorders affecting the vasculature.

Single nucleotide polymorphisms

An important new tool that is expected to contribute in the future to the deciphering of common complex vascular disorders is the use of single nucleotide polymorphisms (SNPs). SNPs are common inherited sequence polymorphisms found frequently (at intervals of roughly every 300–500 bp) throughout the human genome. It has been hypothesized^{12–14} that a subset of these common genetic variants may contribute significantly to the genetic risk of common diseases, such as those affecting the vasculature. Accordingly, the analysis of inter-individual sequence differences in and surrounding genes already known to participate in vascular biology is another area where high-throughput genomic sequencing is being applied to the analysis of disease susceptibility. Several of the initial large-scale studies of sequence variation have focused on genes of relevance to hypertension and atherosclerosis. The examination of 9.7 kb of the lipoprotein lipase gene¹⁵ in 71 individuals revealed 88 variations within the resequenced interval, including 79 single nucleotide variations. Grouping the SNPs on individual chromosomes revealed 88 distinct haplotypes. (A haplotype represents a clustering of the same polymorphic sites shared in chromosomes from many individuals indicating that the segment of DNA in the different individuals has descended from a common ancestral chromosome). In a similar study of particular relevance to hypertension research, the complete



genomic sequence of the angiotensin-converting enzyme was surveyed in 11 individuals¹⁶. Sequencing the entire gene (24 kb) in these individuals identified 78 varying sites that were resolved into 13 distinct haplotypes. The increasing throughput of sequence polymorphism detection resulting from technological advances was illustrated in a much larger study of hypertension where resequencing 'chips' were used. In this study, 'chips' containing 75 hypertension candidate genes (190 kb of sequence) were used to examine 74 individuals¹⁷ identifying almost 900 SNPs. The scheme for the chip-based identification of SNPs is illustrated in Fig. 2.

The sequence polymorphism surveys described above, although providing few biological insights themselves, indicate the feasibility of identifying large numbers of potential SNPs associated with vascular disease and the capability to group these polymorphisms into distinct haplotypes. The burden of haplotype identification compounded by the abundance of polymorphic sites in the human genome make it an arduous task to identify the founder mutations present within the common haplotypes that are hypothesized to contribute to susceptibility to vascular disease. Because these ancient functional polymorphisms probably have small quantitative effects on disease susceptibility, they present a significant challenge in mapping and functional verification. Results from SNP association studies will need to be combined with other approaches, such as the use of genetically tractable animals, to fine map and characterize candidate genes and genetic intervals. The use of SNPs for deciphering human disease is a young discipline and only with time, following extensive application of this approach to disease states, will it be possible to assess its utility in the identification of genes participating in common disorders of the vasculature.

Comparative mapping of complex traits

As a platform to partially overcome the intrinsic difficulties of mapping a complex trait in humans, genome-wide mapping studies using short tandem-repeat polymorphisms have been used extensively to map genes participating in quantitative vascular disorders in mice and rats. The ability to cross genetically well defined, inbred strains that differ in disease susceptibility has enabled the mapping of genes that modify complex phenotypes in rodents in a manner not feasible in humans. In the stroke-prone spontaneously hypertensive rat (SHR), several loci have been mapped and subsequently isolated in congenic strains (animals in which a small chromosomal segment from one strain has been placed through breeding on the background of a second strain), indicating the ability to isolate different loci contributing to this complex quantitative phenotype^{18,19}. Regions of the human genome that are orthologous to many of these hypertension-associated regions in the rat have since been implicated in

various forms of human hypertension²⁰. The identification of atherosclerosis susceptibility genes by crossing resistant and susceptible strains of mice has also been an active field leading to the mapping of multiple susceptibility loci, that is, atherosclerosis susceptibility QTLs²¹.

One difficulty in studying naturally occurring atherosclerosis susceptibility in mice is the subtlety of phenotypic differences between inbred laboratory strains, reflecting the natural resistance of these mice to dietary atherosclerosis. During the past few years the mouse system has begun to be optimized for studying atherosclerosis through the use of mice containing targeted null mutations in the apolipoprotein E (apoE)^{22,23} and the low-density lipoprotein (LDL)-receptor genes²⁴. These animals have increased levels of atherogenic lipoproteins and develop atherosclerotic lesions, mimicking in many ways the lesions seen in humans. Breeding of the apoE null allele into mouse backgrounds that varied in atherogenesis susceptibility resulted in pronounced variability in lesions²⁵. None of the variation in atherosclerosis was attributable to variation in known risk factors such as plasma lipoprotein levels or indices of lipoprotein oxidation, which indicates the participation of new genes for atherosclerosis susceptibility. Insights from quantitative trait mapping in rodents and the application of technologies for manipulating the mouse genome, used in combination with data from human sequence polymorphism studies, should synergistically assist in the identification of genes contributing to common human vascular disorders.

Genome-wide mutagenesis

An important approach for elucidating gene function *in vivo* is the application of genome-wide mutagenesis. Chemical and insertional mutagenesis, which has been used extensively in studies with *Drosophila* and *Caenorhabditis elegans*, offers a robust means to generate new phenotypes, linking genes to function on a genome-wide scale. In response to the widening vertebrate phenotype gap (the difference between the numbers of genes identified and those for which a function has been assigned), chemical mutagenesis is increasingly being exploited to generate new mutant phenotypes in zebrafish²⁶ and in mice²⁷. The overall strategy of genome-wide mutagenesis is illustrated in Fig. 3.

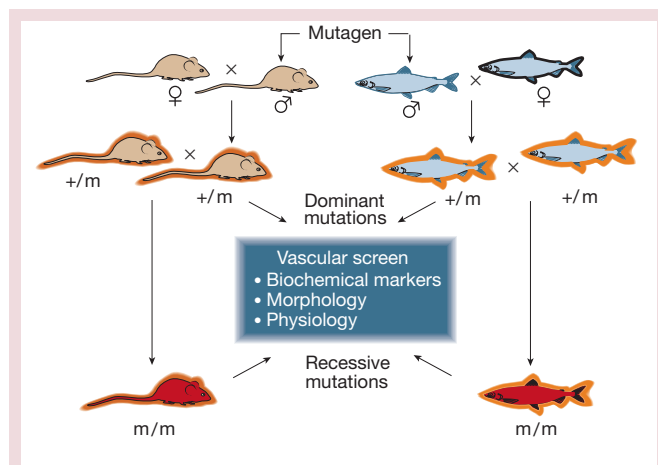


Figure 3 Genome-wide mutagenesis. Phenotype driven mutagenesis is being applied on a large scale in zebrafish and mice to produce mutations affecting vascular development and function. To produce mutant animals, adult males are treated with a chemical mutagen, usually ethylmethanesulphonate, which induces mutations at high frequency throughout the genome. Mutagenized males are then mated to wild-type females. The first-generation progeny of this cross can be screened for the effects of dominant mutations. Alternatively, each first-generation male can be mated to wild-type females, and then mated to the female progeny of this cross to generate animals homozygous for recessive mutations. Phenotypic assays for vascular screens should be specific and high-throughput and may include biochemical, physiological or morphological parameters.

Although the zebrafish is much less studied than the mouse, it has a short generation time (36 days) and significant fecundity (>100 embryos), and is the vertebrate of choice for examining many aspects of vascular development^{28–30}. Within the first 72 hours of fertilization, the zebrafish embryo develops essentially all of the organs specific to vertebrates, most of which are directly visible through the embryo's relatively transparent skin. In addition, the zebrafish is particularly well suited for vascular studies because as an embryo it develops normally up to 7 days past conception even in the absence of a functional heart. As a result of these features, several large-scale zebrafish mutagenesis screens have been carried out, specifically targeting the vasculature.

One of the more informative zebrafish vascular mutations identified in such a screen is *gridlock*³¹. Homozygous *gridlock* embryos have no circulation in their posterior trunk and tail because of failure of the proximal and distal portions of the dorsal aorta to join. The *gridlock* mutation mimics in many ways the clinically important but poorly understood human condition, coarctation of the aorta. Following the identification of a fish embryo with this phenotype, meiotic mapping, genomic sequencing and *in vivo* complementation were used to identify and then prove that a mutation in *grl*, a basic helix–loop–helix protein with a human orthologue, was responsible for the vascular abnormality³⁰.

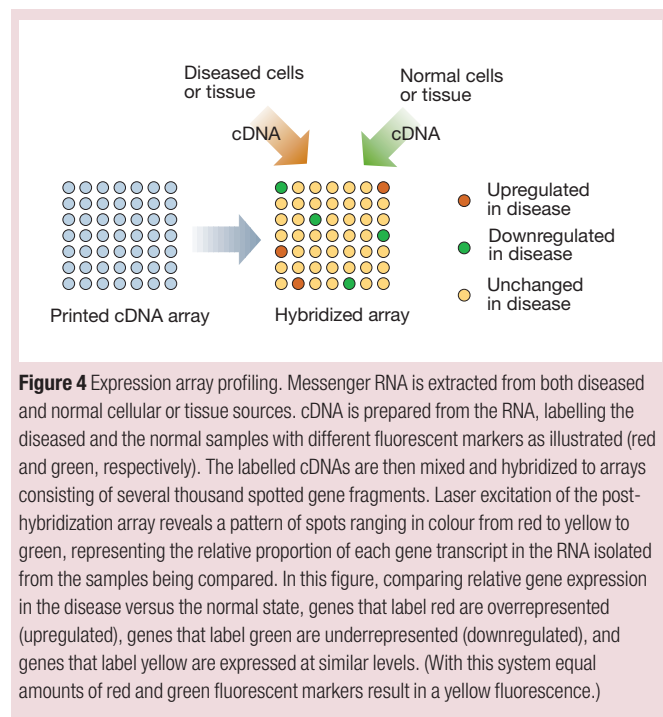
The mouse, a mammal with a well-developed genetic system, is also emerging as a target for genome-wide mutagenesis. Several research centres have begun scaling up the size of murine mutagenesis screens in an attempt to increase the numbers of mutant mouse phenotypes to serve as substrates for linking sequence to function. The Medical Research Council's Mammalian Genetics Unit³² in Harwell, UK, and the Institute of Mammalian Genetics³³ in Munich, Germany, are currently screening several thousand first-generation progeny of chemically mutagenized founders, using a battery of phenotypic assays. These screens have identified hundreds of mutants including many with lipid and cardiac abnormalities, all of which are available to the scientific community for more in-depth analysis.

Genome-wide expression profiling

Complementing the increased availability of sequence, particularly expressed sequences (complementary DNAs), analysis of transcriptional control mechanisms is shifting from predominantly experimental- to information-based methods. In a short period of time, DNA microarrays and other related technologies have matured into robust tools for assessing the expression of an organism's full repertoire of genes in a single experiment³⁴. Expression profiling has already been applied to a variety of issues relevant to vascular biology, including the identification of genes contributing to both single-gene and genetically complex vascular conditions. The general strategy for performing genome-wide expression profiling is illustrated in Fig. 4.

Identification of disease genes

The utility of expression profiling in the identification of genes contributing to vascular disorders is demonstrated in recent studies of Tangier disease in humans³⁵ and syndrome X in rats³⁶. Tangier disease, a rare, recessive genetic disorder mapping to chromosome 9q31 (ref. 37), is associated with the nearly complete absence of circulating high-density lipoprotein (HDL), cholesteryl ester accumulation in tissue macrophages, and heightened atherosclerotic susceptibility. Among the several groups who identified the Tangier disease gene^{9–11}, one group's identification of the gene³⁵ relied on expression profiling and the observation that cells from affected individuals were unable to deliver intracellular cholesterol to nascent HDL particles. Assuming that a defective protein in this pathway was the likely explanation of the clinical finding, fibroblasts obtained from normal and Tangier patients were studied to identify differentially expressed transcripts. Of the 58,800 human cDNAs probed by expression arrays, 160 cDNAs were found to be underexpressed at least 2.1-fold in patients with the disease. From this large number of candidates based on expression changes, only one of them, the *ABCI* transporter, mapped



to a region consistent with the chromosomal assignment of the Tangier defect. Mutations in the *ABCI* gene were shown to be responsible for Tangier disease and are suspected of possibly contributing to other forms of reduced plasma HDL³⁸.

Investigators studying the SHR strain, a model for syndrome X, used a strategy similar to that employed in the search for the Tangier gene to identify a gene contributing to this animals' insulin resistance³⁶. Syndrome X is a complex polygenic trait that includes type 2 diabetes, hypertension, obesity and combined hyperlipidaemia³⁹. Several genetic loci believed to contribute quantitatively to features of this condition have been mapped in the SHR strain. Of these, a single locus on distal rat chromosome 4 was identified as contributing to several aspects of the condition, including insulin resistance. Extensive breeding between wild-type and SHR rats resulted in the generation of congenic animals where only part of chromosome 4 from a hypertensive SHR strain had been replaced with the corresponding region from a non-hypertensive strain. This congenic strain showed partial correction of the insulin-resistance syndrome. Comparisons of adipocyte gene expression between the SHR and the chromosome-4 congenic strain revealed that among the 70 genes that showed at least a twofold change in expression in array systems used, only one, *Cd36*, mapped to the correct chromosomal location. Further analysis revealed that, in the strain of SHR rats used in this study, the *Cd36* gene was rearranged, and the *Cd36* gene product was absent in adipocytes, supporting its participation in the insulin-resistance phenotype.

Genetic mapping, particularly for quantitative vascular disorders, lacks the fine resolution required to pinpoint individual genes. For the studies of both Tangier's disease and rat syndrome X, the linking of information from different genomic resources (mapping and expression profiling) was key to the successful identification of disease genes. Although in many cases this strategy will not be as straightforward as in the two examples described above, the sequencing of the human genome and the chromosomal localization of the sequenced genes should facilitate the identification of defects contributing to most simple, and potentially some genetically complex, disorders affecting the vascular system.

Analysing disease pathways

Genome-wide expression analyses have also been useful in providing an unbiased assessment of an organism's expression response to

common vascular disorders, such as cardiac hypertrophy and myocardial infarction. To seek out genes and pathways whose expression is altered at different stages in the process of cardiac hypertrophy, investigators screened several thousand genes for changes during the induction and regression of hypertrophy in a mouse model⁴⁰. Although these studies were unable to determine which genes have a triggering role in this process, they were able to reveal that distinctly different sets of genes experienced expression changes during induction and regression of hypertrophy. In addition, the finding that a significant percentage of previously uncharacterized new genes showed altered expression during different stages of hypertrophy suggests that many of the genes associated with cardiac hypertrophy have not yet been identified by the more common candidate-gene approaches. A somewhat parallel study examining a rat model of myocardial infarction recently generated a broad picture of the overall patterns of gene expression in different regions of the rat heart during this cardiac event⁴¹. Of the 4,000 genes assessed, the identification of more than 200 genes with an expression response to infarction is clearly greater than that which can be investigated on an individual basis. By emphasizing the clustering of gene expression patterns into functional groups, the investigators were able to derive insights into large-scale processes that the heart uses in response to acute damage and the subsequent functional deficiency.

The validity of expression array results is supported by the many changes that matched those previously characterized in traditional gene-by-gene analyses. A further encouraging aspect of the expression profiling approach for deriving new insights into hypertrophy and myocardial infarction has come from the identification of a significant number of new genes shown to experience altered expression at different stages of these already well-studied processes. As unbiased, genome-wide expression data for a variety of vascular processes increases, and means for statistically analysing these data improves, the emphasis will shift from the identification of individual genes to the development of biological models that fit these data.

Genomics and vascular biology — trends and predictions

Sequence comparisons of the complete genomes of various simple eukaryotes is fast becoming a mature field^{42,43} and very shortly it will be possible to make similar comparisons between the completely sequenced genomes of humans and mice. Advances in sequencing that have already accelerated the completion of the sequencing phase of the human genome programme are unlikely to stop here. It is reasonable to expect that in the distant but foreseeable future we will be able to access the complete, or at least mostly complete, genomic sequence of a large number of different vertebrate species as well as that of individual humans. The availability and analysis of these sequences will contribute to defining the large but finite number of human haplotypes in each segment of an individual's genome, identifying the full set of genes in multiple mammals, and unravelling the non-coding regulatory circuitry controlling the expression of genes participating in vascular biology.

Deciphering the genetics of common but genetically complex vascular disorders is the principal challenge now confronting cardiovascular researchers. It is feasible to expect that large data sets of individual sequence variations, coupled with detailed phenotypic information coming from surveys of an individual's complete repertoire of transcripts and proteins, should contribute in the next few decades to an understanding of the genetic basis of multigenic vascular disorders, paralleling the progress that has already occurred for single-gene disorders. One requirement for the successful application of genomics to vascular biology, and medicine in general, will be the willingness of society to accept and use this information. This will demand an informed population and the protection of individual rights. Assuming this occurs we should see marked improvements in the future in the ability to stratify patients into clinically useful risk and treatment groups and the development of new and customized treatments for diseases affecting the vasculature. □

1. Sanger, F., Nickeln, S. & Coulson, A. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
2. Maxam, A. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl Acad. Sci. USA* **74**, 560–564 (1977).
3. Loots, G. G. *et al.* Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288**, 136–140 (2000).
4. Göttgens, B. *et al.* Analysis of vertebrate *SCL* loci identifies conserved enhancers. *Nature Biotechnol.* **18**, 181–186 (2000).
5. Mayor, C. *et al.* VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* (in the press).
6. Porcher, C. *et al.* The T cell leukemia oncoprotein *SCL/tal-1* is essential for development of all hematopoietic lineages. *Cell* **86**, 47–57 (1996).
7. Robb, L. *et al.* The *scl* gene product is required for the generation of all hematopoietic lineages in the adult mouse. *EMBO J.* **15**, 4123–4129 (1996).
8. Lifton, R. P. Molecular genetics of human blood pressure variation. *Science* **272**, 676–680 (1996).
9. Bodzioch, M. *et al.* The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nature Genet.* **22**, 347–351 (1999).
10. Brooks-Wilson, A. *et al.* Mutations in *ABCI* in Tangier disease and familial high-density lipoprotein deficiency. *Nature Genet.* **22**, 336–345 (1999).
11. Rust, S. *et al.* Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nature Genet.* **22**, 352–355 (1999).
12. Collins, F. S., Guyer, M. S. & Charkravarti, A. Variations on a theme: cataloging human DNA sequence variation. *Science* **278**, 580–581 (1997).
13. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
14. Lander, E. S. The new genomics: global views of biology. *Science* **274**, 536–539 (1996).
15. Nickerson, D. A. *et al.* DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nature Genet.* **19**, 233–240 (1998).
16. Rieder, M. J., Taylor, S. L., Clark, A. G. & Nickerson, D. A. Sequence variation in the human angiotensin converting enzyme. *Nature Genet.* **22**, 59–62 (1999).
17. Halushka, M. K. *et al.* Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nature Genet.* **22**, 239–247 (1999).
18. Kreutz, R. *et al.* Dissection of a quantitative trait locus for genetic hypertension on rat chromosome 10. *Proc. Natl Acad. Sci. USA* **92**, 8778–8782 (1995).
19. Jacob, H. *et al.* Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell* **67**, 213–224 (1991).
20. Stoll, M. *et al.* New target regions for human hypertension via comparative genomics. *Genome Res.* **10**, 473–482 (2000).
21. Mu, J. L. *et al.* Quantitative trait loci analysis for the differences in susceptibility to atherosclerosis and diabetes between inbred mouse strains C57BL/6J and C57BLKS/J. *J. Lipid Res.* **40**, 1328–1335 (1999).
22. Zhang, S. H. *et al.* Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E. *Science* **258**, 468–471 (1992).
23. Plump, A. S. *et al.* Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by homologous recombination in ES cells. *Cell* **71**, 343–353 (1992).
24. Ishibashi, S. *et al.* Massive xanthomatosis and atherosclerosis in cholesterol-fed low density lipoprotein receptor-negative mice. *J. Clin. Invest.* **93**, 1885–1893 (1994).
25. Dansky, H. *et al.* Genetic background determines the extent of atherosclerosis in apoE-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **19**, 1960–1968 (1999).
26. Amsterdam, A. *et al.* A large-scale insertional mutagenesis screen in zebrafish. *Genes Dev.* **13**, 2713–2724 (1999).
27. Vitaterna, M. H. *et al.* Mutagenesis and mapping of a mouse gene, *Clock*, essential for circadian behavior. *Science* **264**, 719–725 (1994).
28. Lee, R. K. *et al.* Cardiovascular development in the zebrafish. II. Endocardial progenitors are sequestered within the heart field. *Development* **120**, 3361–3366 (1994).
29. Stainier, D. Y., Lee, R. K. & Fishman, M. C. Cardiovascular development in the zebrafish. I. Myocardial fate map and heart tube formation. *Development* **119**, 31–40 (1993).
30. Zhong, T. P. *et al.* *gridlock*, an HLH gene required for assembly of the aorta in zebrafish. *Science* **287**, 1820–1824 (2000).
31. Weinstein, B. M. *et al.* *Gridlock*, a localized heritable vascular patterning defect in the zebrafish. *Nature Med.* **1**, 1143–1147 (1995).
32. Nolan, P. *et al.* A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nature Genet.* **25**, 440–443 (2000).
33. deAngelis, M. *et al.* Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nature Genet.* **25**, 444–447 (2000).
34. The chipping forecast. *Nature Genet.* **21**(Suppl.), 1–60 (1999).
35. Lawn, R. M. *et al.* The Tangier disease gene product ABC1 controls the cellular apolipoprotein-mediated lipid removal pathway. *J. Clin. Invest.* **104**, R25–R31 (1999).
36. Aitman, T. J. *et al.* Identification of *Cd36* (*Fat*) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nature Genet.* **21**, 76–83 (1999).
37. Rust, S. *et al.* Assignment of Tangier disease to chromosome 9q31 by a graphical linkage exclusion strategy. *Nature Genet.* **20**, 96–98 (1998). [Published erratum appears in *Nature Genet.* **20**, 312 (1998).]
38. Hayden, M. R. *et al.* Cholesterol efflux regulatory protein, Tangier disease and familial high-density lipoprotein deficiency. *Curr. Opin. Lipidol.* **11**, 117–122 (2000).
39. Reaven, G. M., Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes* **37**, 1595–1607 (1988).
40. Friddle, C. *et al.* Expression profiling reveals distinct sets of genes altered during induction and regression of cardiac hypertrophy. *Proc. Natl Acad. Sci. USA* **97**, 6745–6750 (2000).
41. Stanton, L. *et al.* Altered patterns of gene expression in response to myocardial infarction. *Circ. Res.* **86**, 919–920 (2000).
42. Gelfand, M. S., Koonin, E. V. & Mironov, A. A. Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. *Nucleic Acids Res.* **28**, 695–705 (2000).
43. Koonin, E. V. & Galperin, M. Y. Prokaryotic genomes: the emerging paradigm of genome-based microbiology. *Curr. Opin. Genet. Dev.* **7**, 757–763 (1997).



Vascular biology: a route to novel cardiovascular drugs

Despite advances in awareness, prevention and treatment, cardiovascular disease (CVD) remains the largest cause of death and disability world-wide. With this view of the future, a major mission for the pharmaceutical industry is to develop new CVD therapies that significantly increase life span and improve quality of life for patients. Building on a solid foundation of cardiovascular drug development, beginning in the 60s with beta-blockers and continuing through the 70s and 80s with calcium antagonists and ACE-inhibitors, AstraZeneca is now firmly focusing on the next therapeutic breakthrough opportunities for patients suffering from CVD.

Atherosclerosis, the degenerative process underlying a major part of CVD, is a complex trait arising from the interaction of multiple susceptibility genes with a range of environmental stimuli. Metabolic dysfunction, such as diabetes and dyslipidaemia, accelerates the growth and impact of atherosclerotic lesions. Remodelling of the atherosclerotic vessel wall results in a wide spectrum of plaque architecture, from large, fibrotic lesions typically causing stable angina, to the angiographically occult, lipid rich plaques, prone to rupture. The consequences of plaque rupture can be prevented by antithrombotic therapies. Recent progress in resolution of protein structures in the coagulation cascade has allowed structure based design of highly selective anticoagulants, and AstraZeneca is committed to be a leading scientific force in the development of effective and safe antithrombotic drugs. However, there is also clear need to intervene in the causes of atherosclerotic lesion progression and instability of occult plaques.

The endothelium is important for regulating vascular tone, haemostasis and transport into the blood vessel. Endothelial dysfunction is an early sign of vascular disease, with consequences for atherosclerotic lesion progression and risk of thrombosis. Lack of sufficiently predictive animal models, disease surrogate markers and the need for large cumbersome clinical trials have hampered clinical evaluation of new anti-atherosclerotic drug concepts. During the last decade, advances in transgenic technology have led to successful development of humanized cardiovascular disease models in the mouse. This has established integrative physiology in transgenic mice as a powerful approach for target validation. Evaluation of new atherosclerosis therapies using noninvasive surrogate markers, such as endothelial dysfunction, may offer a simplified bridge between transgenic disease models and rapid concept studies in atherosclerotic patients.

There is now growing interest in therapeutic angiogenesis to induce collateral vessel growth in peripheral and myocardial ischaemic diseases. The contribution of neoangiogenesis in controlling growth and dissemination of solid tumours has led to successful attempts to treat experimental tumours with antiangiogenic approaches. Understanding the fundamentals of angiogenesis is of relevance for therapeutic management of both cardiovascular and cancer diseases.

In the future, treatment of cardiovascular diseases will include tailoring of therapy to the individual, according to a variety of metabolic and genetic profiles. AstraZeneca is therefore participating in extensive genetics programs to identify CVD susceptibility loci and to establish a molecular disease management. Our sponsorship of this Nature Insight reflects the high priority we attach to scientific excellence as the foundation to drug discovery. We look forward to seeing the outcome of advances in vascular biology contributing to improved future therapeutic opportunities for patients suffering from cardiovascular diseases.

Tommy Abrahamsson, Adriano Henney, German Camejo and Mikael Dohlsten
AstraZeneca R & D, Research Area Cardiovascular & Gastrointestinal,
Mölndal, Sweden, and Alderley Park, U.K.



Life in the Fas lane

A selection of reagents and reagent-handling devices.

Greenhouse Parallel Synthesizer

Radleys www.radleys.com

Convenient synthesis in parallel

The GreenHouse 24-position parallel synthesizer was introduced at Achema this year by Radleys Discovery Technologies, a company specializing in products for parallel chemistry. The GreenHouse converts a standard single-place-stirring hotplate into a 24-position stirred, heated and chilled parallel synthesizer with efficient refluxing. It features a removable reaction block holding up to 24 reaction tubes with a reaction volume of 0.5–3 ml. The synthesizer is compact, with a footprint about the same as a standard microtitre plate.

Reader Service No. 100



Parallel line: Radleys' GreenHouse synthesizer.

ARC Catalogue 2000

ARC www.arc-inc.com

Radiochemicals on line

The latest catalogue from American Radio-labeled Chemicals, Inc. lists more than 1,600 products, including over 160 new entries. A standard range of ^{14}C and ^3H labelled products is listed, and ARC also offers custom synthesis. Classifications listed in the catalogue include carbohydrates, nucleotides, lipids, peptides and neurochemical ligands. The range of fatty acids has been expanded to cover 50 fatty acids and coenzyme A derivatives. The web site, arc-inc.com, lists the full catalogue on-line alphabetically and by catalogue number, and features on-line ordering.

Reader Service No. 101

Lysis reagent

Mallinckrodt Baker www.jtbaker.nl

Improved lysing reagent for flow cytometry

The Mallinckrodt Baker Stable Lysing Reagent ultra-stable reagent is both formaldehyde and aldehyde derivative free, generating an environmentally friendly and easy-to-dispose-of product. This also makes it safer for staff working with it. The reagent is supplied in 0.5-litre ready-to-use bottles and has a stable shelf life of at least 18 months. In trials, the manufacturers found that samples lysed with this new reagent showed stable results, even at extended time intervals (10–20 minutes).

Reader Service No. 102

Simplicity Water System

Millipore www.millipore.com

Ultrapure water on tap on the bench-top

This new water purification system is intended for users who need small volumes (under 5 litres a day) of Type 1 water for preparation of

standards and dilution of samples. The Simplicity system produces water at a flow rate up to 0.6 litres per minute. The water produced is free of inorganic ions, organic ions and bacteria, meeting Type I specifications with a resistivity of 18.2 M Ω cm and TOC content of less than 5 p.p.b. The whole set-up has a small footprint (25 × 26 × 46 cm) and can be installed without the need for any plumbing or special mounting. A removable reservoir holds 3.5 litres of pretreated water.

Reader Service No. 103

Neptune Water System

Purite www.purite.com

Delivering ultrapure water

The Neptune system produces ultrapure water at a guaranteed resistivity level of 18.2 M Ω cm. Three dispense modes are available: 'manual dispense' allows the user to simply press the dispense button and hold to obtain the required volume; 'preset volume dispense' can be programmed for the exact volume of water required; and with 'automatic dispense' the user can position a remote sensor in a large container and leave Neptune to fill it unsupervised. A dispense rate of 2 litres per minute is possible on the Analytical setting.

Reader Service No. 104

Vectastain Reagents

Vector Laboratories www.vectorlabs.com

Prediluted histochemical staining kits

The Vectastain Universal Elite ABC Kit and the Universal Quick Kit are now available in stable, prediluted, ready-to-use (RTU) formats. These products are recommended for

automated immunohistochemical staining systems or traditional manual protocols and provide the same sensitivity as their concentrated counterparts, in a more convenient form. The RTU Universal Elite ABC Kit contains prediluted blocking serum, diluted biotinylated universal secondary antibody, and the RTU Elite ABC Reagent (also available separately). The RTU Universal Quick Kit contains prediluted blocking serum, diluted biotinylated pan-specific secondary antibody, and the RTU streptavidin/peroxidase complex. All reagents are stable for more than 12 months. Each kit contains reagents to stain approximately 500 sections.

Reader Service No. 105

Flask Carrier Cart

Labconco www.labconco.com

A wheel advance

Labconco's Flask Carrier Cart holds twenty-four 500 or 800 ml volumetric or Kjeldahl flasks either upright or neck-down for easy draining. A removable drip pan minimizes clean-up, and vinyl coated steel neck supports hold upright flasks securely. The chassis frame is made from one-inch square steel tubing with corrosion-resistant baked-on epoxy powder coating, charcoal in colour. The four-inch hard rubber casters stabilize movement over uneven floors. Top and bottom shelves have melamine high-pressure laminate surfaces and bevelled holes, and are made of 10-mm-thick fibreglass board. The Flask Carrier Cart from Labconco is delivered ready to roll with no assembly required.

Reader Service No. 106



Cart your flasks around.

new on the market

Acetone RTK Safety Wash Bottle

Nalgene www.nalgenunc.com

Have you got the bottle?

Nalgene's latest 500 ml Acetone RTK Safety Wash Bottle is made of LDPE (low density polyethylene) by a new process that provides LDPE with enhanced chemical resistance. The new bottle also features a non-metallic self-vent to reduce 'sputter', keeping the contents in the bottle and eliminating the risk of solvent leakage. The manufacturing process also provides a softer texture to make the bottle easier to squeeze. The new bottle is available separately or as part of an assortment containing bottles to hold isopropanol, methanol, sodium hypochlorite (bleach) and distilled water.

Reader Service No. 107

LucLite Plus

Packard www.packardinst.com

Detect firefly luciferase down to 8 fg

The LucLite Plus reagent kit is designed to simplify luciferase reporter gene assays in higher density microplate formats. The kit provides a luminescent signal that is 250% brighter than LucLite. Good mixing, media compatibility and sensitivity make the LucLite Plus kit suitable for 384-well microplate-based reporter gene assays. Combined with liquid handling

systems such as Packard's MultiPROBE II or PlateTRAK, or the TopCounts Microplate scintillation and luminescence counter, LucLite or LucLite Plus form a complete HTS platform for luciferase reporter gene assays.

Reader Service No. 108

FlexiChem Brochure

Robbins Scientific www.robsci.com

An upgrading for this synthesis system

The FlexChem Organic Synthesis System now features a 48-well synthesis reactor with a 5.5-ml chamber volume. This new reactor can be used for larger-scale syntheses using over 300 mg resin per chamber. Also recently released is a new chemically resistant, multi-channel dispenser for adding solvents or reagents to FlexChem reactors. The FlexChem Solvent Dispenser replaces expensive, hand-held, plastic multi-channel pipettors that can be damaged by many commonly used solvents.

Reader Service No. 109

SuperFas Ligand

Alexis Biochemicals www.alexis-corp.com

For Fas signalling pathway studies

This highly active SuperFas Ligand is a recombinant human protein produced in human embryo kidney cells. It kills Fas-sensitive cells

at under 1 ng per ml. Fas ligand (FasL) is a 40kD type II transmembrane protein belonging to the TNF family. FasL is implicated in CTL-mediated killing, activation-induced cell death, creation of immune-privileged sites and tissue homeostasis. Circulating sFasL can be detected in serum of patients with melanoma and some lymphomas.

Reader Service No. 110

Cytospray Plus

Streck Laboratories www.streck.com

A 'new generation' cytology fixative

CytoSpray Plus uses a true fixative in an alcohol base to stabilize and fix cells to a slide. The product is claimed to provide cellular protection and enhanced morphology with crisper detail and less background for Pap smears than alternative methods. It reduces desiccation and specimen loss of cervical samples. Research applications include ICC, FISH, *in situ* PCR and RT-PCR for the detection of HPV or HIV. CytoSpray Plus is packaged in 12 × 4 ounce or 24 × 1 ounce cases. The product can be stored at room temperature, with a closed-bottle stability of 18 months.

Reader Service No. 111

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

Diagnostic Reagents

Pefachrome® Series: chromogenic and fluorogenic substrates

Pefabloc® Series: serine protease inhibitors

Purified snake venom proteins (e.g. Protac®, Ecarin, Batroxobin)

PENTAPHARM LTD
CH 4002 Basel Fax: +41 61 310 00 10
<http://www.pentapharm.ch>

READER ENQUIRY NO. 16

L-012 SODIUM SALT
CHEMILUMINESCENCE PROBE

- CHL Intensity up to 100 times greater than MCLA & Luminol
- Works on cells & whole blood
- Ideal for Precious Cells & High Throughput Assays

Wako Wako USA 877-714-1921 WAKO GmbH (49)02131 3110
www.wako-chem.co.jp

READER ENQUIRY NO. 34

Protein & Peptide Sequencing Service
by **Biologica Co.**

N-terminus Sequencing including in-gel digestion, Microfractionation

You only send us Gel spot or piece from 2D-E or PAGE containing a few pmol protein.

bo For Health and Science **Biologica co.**
E-mail: biologic@rr.iij4u.or.jp
<http://www.iijnet.or.jp/biologica/>

READER ENQUIRY NO. 49

BACs • YACs • PACs
cDNAs

Research Generation
11000 University Avenue, Suite 400
La Jolla, CA 92037
Tel: (619) 594-1300
Fax: (619) 594-0000

Check out our homepage at
<http://www.genegen.com>

READER ENQUIRY NO. 40

New Visions in Digital Imaging

Alpha Imaging Corporation

READER ENQUIRY NO. 48

Look no further

Oligo Synthesis

At GenScript, we provide the highest quality oligos for your research. We offer a wide range of services including:

- Custom oligo synthesis
- Oligo pools
- Oligo libraries
- Oligo arrays
- Oligo conjugates
- Oligo modifications

For more information, please contact us at:

GenScript, Inc. 1500 Shattuck Avenue, Suite 100
Berkeley, CA 94709
Tel: (415) 872-2000
Fax: (415) 872-2001
Email: info@gen-script.com
Web: <http://www.gen-script.com>

READER ENQUIRY NO. 33